Facilitating adults' initial L2 syntax acquisition

The role of cross-linguistic syntactic similarity and lexical processing

Noèlia Sanahuja Cobacho

Universidad del País Vasco
Euskal Herriko Unibertsitatea
Master eta Doktorego Eskola
Escuela de Máster y Doctorado
Master and Doctoral School

Supervisor: Dr. Kepa Erdocia

2024

Abstract

Learning a second language is challenging and L2 learners rarely achieve native-like proficiency (R. Ellis, 2004; Housen & Simoens, 2016). In this dissertation, I investigate how to facilitate adults' initial L2 syntax acquisition, focusing on the role of cross-linguistic syntactic similarity and lexical processing. In five behavioural experiments, I explore how complete beginner L2 learners learn cross-linguistically similar vs. dissimilar L2 structures (Experiment 1 and Experiment 2) and how the acquisition of these structures is influenced by word frequency (Experiment 2 vs. Experiment 3) and cognateness (Experiment 4 and Experiment 5). On the one hand, several theories predict how L1 syntax may influence the acquisition of similar and dissimilar L2 structures (e.g. MacWhinney, 2005; B. D. Schwartz & Sprouse, 1996; Westergaard, 2021). Yet, whether cross-linguistic syntactic similarity eases syntax acquisition by novice adult L2 learners is under-researched. On the other hand, high frequency words and cognates, which are more strongly activated than low frequency words and non-cognates, might facilitate L1 and L2 syntax processing (Hopp, 2016, 2017). However, whether these words facilitate L2 syntax acquisition has not been studied and, to my knowledge, no theory addresses this. The MOGUL framework (Sharwood Smith & Truscott, 2014) comprehensively accounts for first and second language processing and acquisition. In this thesis, I elaborate on how MOGUL could postulate hypotheses about the facilitative role of cross-linguistic syntactic similarity and lexical processing in initial L2 syntax acquisition and I provide experimental evidence to support them.

In Chapter 1, I provide an overview of MOGUL and how I propose it might address the topics studied in this thesis. I also detail the aims and outline of the dissertation. In Chapter 2, I show that Spanish natives with no knowledge of Galician demonstrate greater learning of a Spanish-Galician similar structure, which can be processed with an L1 structure, than of a Spanish-Galician dissimilar structure, which must be acquired from input. In Chapter 3, I investigate whether learning of these structures is affected by lexical frequency. In MOGUL, the higher the activation of a word is during processing, the higher the activation of the structure containing it and the greater the learning of the structure resulting from processing. In line with this, I show that greater learning is achieved when the L1-L2 dissimilar structure is processed with high frequency words than with low frequency words. By contrast, this facilitation is not observed for the L1-L2 similar structure, which strictly speaking does not need to be learnt. In Chapter 4, I study the influence of cognates in acquiring cross-linguistically similar and dissimilar L2 structures. Spanish natives without knowledge of Basque learn a mini-language with Basque lexicon and structures either similar or dissimilar to Spanish. Mirroring Chapter 3, I find that the stronger activation of cognates compared to non-cognates eases the acquisition of L1-L2 dissimilar structures, but not of L1-L2 similar structures. Finally, in Chapter 5 I summarise the findings and conclusions of the thesis. In sum, this dissertation broadens our understanding of the facilitation exerted by cross-linguistic syntactic similarity and lexical processing at the earliest stage of adults' L2 syntax acquisition.

Acknowledgements

I would like to express my sincere gratitude to a large number of people who have helped me during the PhD journey. First, I would like to thank my supervisor, Dr. Kepa Erdocia, for being so supportive, kind and for helping me grow into an independent researcher. Second, I would like to thank the members of the *Gogo Elebiduna* research group. Special thanks must go to Yolanda, for helping me program the experiments in this thesis, and to all the predoctoral researchers who have shared these years with me and have helped me not only providing valuable insights into my research topic, but also valuable emotional support. Next, I would like to thank my supervisors at the University of Barcelona and at the Technical University of Brunswick, where I conducted research stays —Dr. Ruth de Diego-Balaguer and Dr. Holger Hopp— for helping me with theoretical and methodological aspects of the thesis. Finally, I would like to thank four people who constitute my essential support system. To my parents and my brother— thank you for always cheering me up and for accepting my decision to continue with this project even when you saw me down, dealing with numerous setbacks and challenges. To Luis— thank you for believing that I could do this even when I did not believe it myself and for always being willing to give a helping hand without asking anything in return. I could not have done this without you.

Contents

List of Figures

List of Tables

List of Abbreviations

3SG	Third person singular
3PL	Third person plural
A	Agent
ADJ	Adjective
ABS	Absolutive
ACC	Accusative
AGR	Agreement
AIT	Autonomous Induction Theory
ANOVA	Analysis of Variance
APT	Acquisition by Processing Theory
BIA	Bilingual Interactive Activation
BIA+	Bilingual Interactive Activation plus
BIA-d	Developmental Bilingual Interactive Activation
C	Cognate
COMP	Complementizer
CORPES XXI	Corpus del Español del Siglo XXI (Corpus of the 21st Century Spanish)
CFE	Cognate Facilitation Effect
CI	Confidence Interval
CL	Clitic
Congr	Congruent
CP	Complementizer Phrase
D	Determiner
DO	Double Object
DP	Determiner Phrase
DS	Dissimilar Structure
ERG	Ergative
ERP	Event-Related Potential
FT/FA	Full Transfer/Full Access
GJT	Grammaticality Judgement Task
HA	High Attachment
Impl	Implausible
INF	Infinitive
INFL	Inflection
Incongr	Incongruent
L1	First/native language
L2	Second language
LA	Low Attachment
LD	Levenshtein Distance

M	Mean
MOGUL	Modular On-line Growth and Use of Language
N	Noun
NA	Not Available
NC	Non-cognate
NOM	Nominative
NP	Noun Phrase
OSV	Object-Subject-Verb
OVS	Object-Verb-Subject
P	Patient
PJT	Plausibility Judgement Task
Pl	Plausible
PO	Prepositional Object
PP	Prepositional Phrase
PREP	Preposition
PRS	Present
PST	Past
PTCP	Participle
RHM	Revised Hierarchical Model
SBJV	Subjunctive
SD	Standard Deviation
SOV	Subject-Object-Verb
SS	Similar Structure
SVO	Subject-Verb-Object
UCM	Unified Competition Model
UG	Universal Grammar
V	Verb
VP	Verb Phrase
VSO	Verb-Subject-Object
WFE	Word Frequency Effect

Chapter 1

General introduction

Learning languages is very relevant nowadays, in a time of globalisation where people around the world are interconnected. Globalisation has led to hypermobility; people frequently travel to other nations and continents for reasons of work, education or tourism. Additionally, numerous technological advances have opened the door to worldwide communication without the need of travelling and have facilitated knowledge and information mobility. Globalisation has also led to an international economy that interrelates countries at a global scale and learning the languages that allow participating in this economy has become essential. In this context, knowing more than one language is highly desirable and, in the last decades, the world has seen an increasing interest in language learning (Pauwels, 2014; Wang, 2023). In linguistics, the learning of a language after the *native* or *first language* (L1) has been acquired is referred to as *second language (L2) acquisition*. L2 acquisition, just as L1 acquisition, requires learning a large number of words (including their meaning and pronunciation) and the syntax of the new language (the ways words are put together to construct sentences), amongst other challenges. Children differ in their speed of acquisition but, except in cases of extreme environmental privation (e.g. Curtiss, 1977), they all achieve full competence in the L1. By contrast, L2 acquisition is usually more difficult. Learners vary in their rate of acquisition and in their ultimate level of competence, with only a few reaching native-like proficiency (R. Ellis, 2004; Housen & Simoens, 2016). Many factors contribute to the variability in L2 acquisition. For instance, age of acquisition (age of arrival/immersion in the L2-speaking community) has been long recognized to affect the outcome of the learning process, in the sense that L2 proficiency usually declines with increasing age of acquisition (e.g. Bialystok & Hakuta, 1999; Birdsong, 1992; Birdsong & Molis, 2001; Flege, 1995)[1]. Other factors affecting L2 proficiency are language aptitude (e.g. Abrahamsson & Hyltenstam, 2008; Granena, 2013), motivation (e.g. Alamer & Alrabai, 2023; Alamer & Lee, 2019), anxiety in the foreign language class (e.g. Alrabai, 2022; C. Li & Wei, 2023), working memory (e.g. Linck & Weiss, 2011; Sagarra, 2017), linguistic distance (Chai & Bao, 2023; Schepens et al., 2013) and more (see S. Li et al., 2022 for an overview).

[1] This does not mean, however, that learners with a late age of acquisition (e.g. after puberty) cannot reach a native-like proficiency in the L2 (see, for instance, Birdsong, 1992; Kinsella & Singleton, 2014; Muñoz & Singleton, 2007).

Throughout this dissertation, I investigate how adult L2 acquisition can be facilitated. My focus is on the *initial stages* of acquisition. That is, I study acquisition from learners' first exposure to the L2. Of the many aspects of language that need to be acquired, I focus on syntax and, specifically, on syntactic structures, capturing the ways in which words are arranged into constituents within the sentence (Van Valin, 2001). In this thesis, I thus study how adults' *initial second language syntax acquisition* can be facilitated. This question could be approached from many perspectives. I could explore the potentially facilitative role of some of the aforementioned learner-related variables (e.g. motivation or working memory). I could also study the possible facilitation exerted by the learning context (e.g. implicit vs. explicit), a topic which has attracted considerable interest in L2 acquisition research (e.g. Alanen, 1995; de Graaff, 1997; DeKeyser, 1995; more recently Tagarelli et al., 2016). In this dissertation, I will explore the facilitation stemming from two linguistic factors: cross-linguistic syntactic similarity and lexical processing.

On the one hand, there is extensive evidence that cross-linguistic influence facilitates learning of structures and morphosyntactic features that are similar in the L1 and the L2, as measured by learners' processing and use of these structures and features at different stages of L2 proficiency (e.g. Chang & Zheng, 2015; Foucart & Frenk-Mestre, 2011; Izquierdo & Collins, 2008; Tokowicz & MacWhinney, 2005). However, whether cross-linguistic similarity additionally facilitates syntax acquisition by complete beginner adult L2 learners remains an under-researched topic. On the other hand, some words are easier to process than others are, depending on factors such as frequency of occurrence, cognateness, concreteness, neighbourhood density or word length, amongst others (see Barber et al., 2013; Carreiras et al., 1997; Dijkstra et al., 1999; Grainger et al., 2005; Kroll & Merves, 1986; Lemhöfer et al., 2008; Lemhöfer & Dijkstra, 2004; Ziegler et al., 2001, and more). For instance, *high frequency words*, i.e. words that are often encountered in the language, are processed faster and more accurately than *low frequency words*, i.e. words that are encountered less often (Cop et al., 2015; Duyck et al., 2008; Gollan et al., 2008, 2011; Whitford & Titone, 2012, and more). A processing advantage is also found for *cognates*, words with similar form and meaning in two or more languages (e.g. English *tomato* – Dutch *tomaat*) over *non-cognates*, words with similar meaning and different form across languages (e.g. English *tomato* – Italian *pomodoro*) (Costa et al., 2000; Dijkstra et al., 1999, 2010; Lemhöfer & Dijkstra, 2004; Van Assche et al., 2011, and more). Previous evidence suggests that high frequency words (Hopp, 2016; Tily et al., 2010) and cognates (e.g. X. Chen et al., 2023; Hopp, 2017) facilitate L1 and L2 syntactic processing. Nevertheless, to the best of my knowledge, whether and how differences in lexical frequency or cognateness affect the acquisition of L2 syntactic structures has not been investigated. In this dissertation, I present evidence that contributes towards narrowing down these gaps in the literature, showing that cross-linguistic syntactic similarity, word frequency and cognateness facilitate adults' initial L2 syntax acquisition. This evidence comes from 5 behavioural experiments where I study how complete novice L2 learners acquire cross-linguistically similar vs. dissimilar L2 structures (see Chapter 2) and how the acquisition of this type of structures is affected by lexical frequency (see Chapter 3) and cognate status (see Chapter 4).

Making detailed hypotheses about how cross-linguistic syntactic similarity and lexical processing could affect initial L2 syntax acquisition requires a broad understanding of the bilingual mind. More specifically, it is necessary to comprehend how the lexicon and the syntax of the L1 and the L2 (including words varying in frequency and cognateness, on the one hand, and cross-linguistically similar and dissimilar structures, on the other hand) are represented, processed and/or acquired, and how lexical and syntactic processing interact in the course of L2 acquisition. As far as I know, there is no model, theory or framework explicitly specifying all these aspects (but several theories provide insights into some of these topics separately, see Section 2). Crucially, there is a theoretical framework which provides a detailed and comprehensive view about how languages are represented, processed and acquired in real time and which, I propose, can accommodate claims from multiple of the aforementioned theories: The Modular On-line Growth and Use of Language (henceforth, MOGUL) framework (Sharwood Smith, 2017; Sharwood Smith & Truscott, 2014). This framework will serve me as a basis to postulate hypotheses and make predictions for my experiments throughout the dissertation.

In what follows, I describe how some terms and concepts related to second language acquisition are understood in this thesis (Section 1). In Section 2, I turn to the idea of facilitating initial L2 syntax acquisition within this dissertation. In this context, I discuss the need for a theoretical approach that allows investigating this topic. I review some existing theories and their shortcomings and I present the MOGUL framework as a suitable option. Next, in Section 3, I provide an overview of this framework. I introduce its architecture, the way it generally conceives language processing and acquisition, the differentiation of the L1 and the L2 in the bilingual mind and the influence of L1 syntax and lexical processing (focusing on high frequency/low frequency words and cognates/non-cognates) on the initial acquisition of L2 syntax. Finally, in Section 4 I present the aims and outline of the following chapters of the dissertation.

1. Important terminology and concepts

In this section, I first define the term *second language*, focusing on the difference oftentimes assumed between (i) *second language* and *foreign language* and (ii) *second language* and *third language*. Then, I describe the distinction between *acquisition* and *learning* and how these two terms are connected to the related notions of *implicit* and *explicit* learning. In my experiments, I used implicit and explicit learning paradigms, so I finally discuss their characteristics and the role they play in my dissertation.

1.1. Second language

Some researchers define the term *second language* attending only to the context in which the language is learnt. A *second language* is learnt in a country where it is used as a medium of communication, either because it is the native language (e.g. English in the United States) or because it is used socially or institutionally (e.g. English in Singapore). In that context, the learner acquires the second language by being exposed to it naturally. This definition of *second language* can be opposed to that of *foreign language,* understood as a language learnt in a country where it is not the L1 of the majority of the population and where it is not usually spoken or used for communication. In that context, the learner acquires the foreign language in the classroom, e.g. learning English in Spain (Miao, 2015). On the other hand, some researchers use the term *second language* to refer to the language learnt immediately after the first language and oppose it to the term *third language,* referring to the language learnt after the L2. Researchers making this distinction are usually interested in the influence of the first or second language on the development of the third one and vice versa (Arıbaş & Cele, 2021; Cherici, 2023; Gut, 2010; Hammarberg, 2001 and many more). In spite of this, most scholars in the field of second language acquisition use *second language* to refer to any language different from the first one, irrespective of the context and the order in which it is learnt (VanPatten & Williams, 2015). This is also how the term *second language* should be interpreted in my thesis.

1.2. Acquisition and learning

Researchers on L2 acquisition sometimes differentiate between *acquisition* and *learning.* This distinction was first proposed in the 1970/80s within Krashen's *Monitor Theory,* the first one developed particularly for second language acquisition (Krashen, 1977, 1981, 1982, 1985). Krashen proposed that *acquisition* denotes a process of obtaining knowledge that occurs unintentionally and without instruction as learners are exposed to *comprehensible input* in the L2 and try to understand it[2]. Acquired knowledge is usually unconscious and not verbalizable. In this sense, L2 acquisition would be very similar to L1 acquisition. By contrast, *learning* is a process of obtaining knowledge that occurs intentionally via explicit instruction. The resulting L2 knowledge is conscious and verbalizable (e.g. in the form of grammatical rules or patterns). Crucially, according to Krashen only acquisition improves the language a speaker knows and only acquired knowledge can be used in spontaneous speech; learnt knowledge is just used to monitor acquired knowledge during production. This claim would explain, for instance, why learners might use a structure correctly without being able to articulate it and why they might verbalize a grammatical pattern of the L2 and not be capable of using it in production. In spite of this, the Monitor Theory has received significant criticism through the years. Regarding the acquisition/learning distinction, VanPatten and Williams (2015) point out that it is difficult to test and confirm that learners use acquired knowledge (and not learnt knowledge) during L2 production, particularly when confronted with evidence that learners

[2] *Comprehensible input*: Language that is somewhat above the level of the language learners know at a given moment.

use the L2 spontaneously and accurately after receiving just formal instruction, which usually includes very little comprehensible input. The terms *acquisition* and *learning* are often used as synonyms in the literature (Rieder-Bünemann, 2012). I will also follow this practice in my thesis.

After the publication of the Monitor theory, researchers on second language acquisition and teaching made a connection between the concepts of *acquisition* and *learning* and *implicit* and *explicit* learning. Krashen's *acquisition* was associated with *implicit learning,* a term coined by Reber (1967) that defines the learning process that occurs without intention to learn and without awareness of what has been learnt. By contrast, what Krashen called *learning* was related to *explicit learning,* referring to a learning process that is intentional and results in learners being aware of their knowledge (Rebuschat, 2013, 2015). Krashen's theory and its implications for L2 teaching boosted the interest in implicit and explicit L2 learning, a topic that is still investigated nowadays.

1.3. Implicit and explicit learning paradigms

Given the complexity of natural languages, in the last decades, most researchers have studied L2 learning under implicit or explicit conditions using *artificial, semi-artificial* or *miniature* (*mini*) *languages*. Artificial languages consist of unknown lexicon and grammatical regularities that often (but not always) mirror natural language grammars (e.g. de Graaff, 1997; DeKeyser, 1995; Morgan-Short et al., 2010, 2014). Semi-artificial languages usually combine vocabulary in the L1 or a highly familiar L2 with syntax from an unfamiliar language (e.g. Alanen, 1995; Rogers et al., 2015; Tagarelli et al., 2016; Williams, 2005). Finally, mini-languages are reduced language versions based on a particular natural language (e.g. Bastarrika & Davidson, 2017; Tolentino & Tokowicz, 2014). In all cases, these languages allow for a rigorous control of the lexicon and the syntax that learners receive and, due to their small size, they can be learnt in a relatively short time. In my dissertation, I use mini-languages. I investigate the acquisition of these mini-languages using, as advanced, both implicit and explicit learning paradigms, so I now define and contextualize them within my research.

Implicit and explicit learning paradigms usually have three parts: (i) an exposure phase, during which learners are exposed to the structures or features to be learnt[3], (ii) a testing phase, which measures learning and (iii) an awareness measure, which determines whether the knowledge learnt, if any, is conscious or unconscious. As Rebuschat (2013) points out, not all experiments using these paradigms include an awareness measure. This is seen as a drawback, since these measures provide useful information about the knowledge acquired under specific learning conditions. All the experiments in my dissertation include an exposure phase, a

[3] In many psycholinguistic experiments, learners are assumed to be learning morphosyntactic rules, which the authors of these experiments express metalinguistically (e.g. Kim & Fenn, 2020; Rebuschat, 2009; Robinson, 1997; Tagarelli et al., 2016; Williams, 2005).

testing phase and an awareness measure. The difference between implicit and explicit learning paradigms resides in the specifics of their parts.

In an implicit learning paradigm, exposure is incidental; learners are not told that they have to learn some regularities nor that they will be tested on their knowledge afterwards (Rebuschat, 2013, 2015). Learners are expected to acquire structures or morphosyntactic features without their attention being devoted to this purpose, i.e. unintentionally. One way to achieve this is to have learners extract formal regularities from the input while performing a meaning-oriented task, for instance a Plausibility Judgement Task (PJT) (e.g. Kim & Fenn, 2020; Maie & DeKeyser, 2020; Rebuschat, 2009, Experiments 1-5; Tagarelli et al., 2016; Williams & Kuribara, 2008). Learning is assessed in the testing phase. One of the most extensively used tests is a Grammaticality Judgement Task (GJT), which asks learners to discriminate between grammatical regularities presented in the exposure phase and ungrammatical regularities, never encountered before (e.g. Kim & Fenn, 2020; Maie & DeKeyser, 2020; Rebuschat, 2009; Tagarelli et al., 2016; Williams & Kuribara, 2008). Learning is considered to have taken place if learners are sensitive to the difference between grammatical and ungrammatical sentences (e.g. if they accept the former and reject the latter). Incidental exposure aims to produce implicit learning. Whether learning has actually been implicit or not is typically inferred from the awareness measure, which assesses the product of this learning (R. Ellis, 2009). The output of implicit learning is unconscious or implicit knowledge. Knowledge of what has been learnt is considered implicit if it is not verbalizable (e.g. Kim & Fenn, 2020; Leung & Williams, 2006; Rebuschat, 2009; Tagarelli et al., 2016; Williams, 2005).

In an explicit learning paradigm, learning is the result of intentional instruction. This encourages learners to consciously think about syntactic patterns during exposure either deductively, as in explicit grammar teaching or inductively, as in a structure-search task (also referred to as rule-search task). Both types of intentional instruction have been widely used in the literature (e.g. Bastarrika & Davidson, 2017; de Graaff, 1997; Kachinske et al., 2015; Rebuschat, 2009; Robinson, 1996; Rosa & O'Neill, 1999; Tagarelli et al., 2011; Tagarelli et al., 2016; Tolentino & Tokowicz, 2014). In addition, learners can also be prompted to reflect on syntactic regularities in the testing phase, through feedback. If the test is a GJT, feedback can provide explicit information about learners' accuracy when judging grammatical and ungrammatical sentences. This can promote a conscious focus on form that makes the target structures or features more salient (Leeman, 2007)[4]. Intentional instruction seeks to result in explicit learning. As mentioned, the type of learning is determined by looking at the output of

[4] Some implicit and explicit learning experiments refer to the task used to assess learning of structures or features as a Grammaticality Judgement Task. Yet, learners may have not been explicitly instructed on the grammar of the L2 and, hence, may not know what is "grammatical" or "not grammatical/ungrammatical". To cope with this, researchers usually instruct learners to judge sentences as being possible/not possible or correct/incorrect according to their knowledge or intuition, sometimes using these terms to define the options *grammatical* and *not grammatical/ungrammatical* and others avoiding the use of these words.

the learning process. Explicit learning usually produces conscious or explicit knowledge. If the awareness measure is a verbal report, explicit syntax knowledge should be verbalizable.

Notably, the relationship between incidental exposure, implicit learning and implicit/ unconscious knowledge, on the one hand, and intentional exposure, explicit learning and explicit/conscious knowledge, on the other hand, is not always as straightforward as described. Incidental exposure may result in explicit learning, if learners figure out the aim of the exposure phase and consciously think about the language presented. Likewise, intentional exposure may result in implicit learning, if learners are not motivated to pay attention to the input and end up learning regularities in the language without intention to (R. Ellis, 2009). It may also be the case that incidental or intentional exposure triggers explicit knowledge for some learners and implicit knowledge for others (e.g. Kim & Fenn, 2020; Robinson, 1997; Williams, 2005).

Explicit and implicit learning paradigms have been widely used to study L2 syntax acquisition (e.g. de Graaff, 1997; 1995; Kim & Fenn, 2020; Leow, 2000; Maie & DeKeyser, 2020; Rebuschat, 2009; Robinson, 1996, 1997; Rosa & O'Neill, 1999; Tagarelli et al., 2016; VanPatten & Oikkenon, 1996; Williams, 2005; Williams & Lovatt, 2003). The focus of these studies has been on whether implicit syntax learning can take place and, if so, for what aspects, whether conscious or explicit knowledge is necessary for L2 syntax learning, whether a paradigm results in implicit or explicit knowledge, the effectiveness of one paradigm compared to the other, and more. Importantly, my experiments were not designed to explore any of these research questions, even if they can answer some of them. Instead, the learning paradigms were used as mere tools to study initial L2 syntax acquisition. In all the experiments in this dissertation, learners were expected to discover regularities in the input on their own (sometimes intentionally, sometimes unintentionally). Thus, I studied learners' capacity to extract knowledge from input, an ability that governs first language acquisition and other human skills, such as perceiving music (Dienes & Longuet-Higgins, 2004; Rohrmeier et al., 2011) and socializing (Lewicki, 1986).

2. Facilitating initial L2 syntax acquisition

As mentioned at the beginning of the chapter, in this thesis I study the facilitative role of cross-linguistic syntactic similarity and lexical processing (of high frequency words vs. low frequency words and cognates vs. non-cognates) in the initial acquisition of L2 structures (similar and dissimilar to L1). Investigating this requires understanding (i) how L1-L2 similar and dissimilar structures are represented, processed and acquired, (ii) how words varying in lexical frequency and cognate status are represented and processed in the bilingual mind and (iii) how syntactic and lexical processing of the aforementioned structures and words interact during real-time L2 acquisition. Explaining all the above is a complex matter and, to facilitate research, scholars have studied some of these topics in isolation.

On the one hand, a wide range of theories and models describe the initial state of L2 acquisition and hypothesize how the L1 may influence the acquisition of L2 structures having or lacking a similar counterpart in the first language. Some examples are the Full Transfer/Full Access model (FT/FA model, B. D. Schwartz & Sprouse, 1994, 1996), the Autonomous Induction Theory (AIT, Carroll, 1999, 2001), the Unified Competition Model (UCM, MacWhinney, 2005) and the Micro-cue model of L2 acquisition (Westergaard, 2021). These accounts differ in more than one aspect. The UCM views language acquisition and processing as the result of mapping linguistic forms or cues (e.g. word order, animacy) to functions or meanings (e.g. agency), with L2 syntax acquisition initially relying on L1 cues. The FT/FA model and the AIT assume that L2 syntax learning occurs when the processing system fails to process L2 input using L1 grammar. By contrast, the Micro-cue model rejects the view that processing can either succeed or fail. It conceives L2 syntax learning as entirely the result of processing, with the processing system initially using L1 micro-cues (structures capturing micro-variation in language) and postulating new micro-cues when needed (see Chapter 2 for more details on these theories). In spite of their differences, these accounts all agree that the L1 is present at the initial state of L2 acquisition and that L1 syntax (including structures, cues or micro-cues) will be used to parse L2 input. When a structure is cross-linguistically similar, processing will be facilitated. When a structure is cross-linguistically dissimilar and, hence, the L2 input cannot be parsed according to L1 syntax, the language system will need to be restructured, a process that is naturally more costly. Importantly, within the theories and models mentioned, the FT/FA model and the Micro-cue model focus just on explaining cross-linguistic influence in L2 syntax learning and do not specify how lexical items are represented and processed in the bilingual mind. Conversely, the AIT and the UCM do discuss lexical representation and processing, but do not mention how words differing in lexical frequency and cognate status would be represented and processed by the L2 learner. None of the theories reviewed discuss the interaction between lexical and syntactic processing during L2 acquisition.

On the other hand, several proposals have been advanced regarding the representation and processing of high vs. low frequency words and cognates vs. non-cognates in the bilingual mental lexicon. A well-known model of word recognition and comprehension accounting for this is the Bilingual Interactive Activation plus (BIA+) model (Dijkstra & van Heuven, 2002). This model assumes that lexical access is language non-selective and occurs in an integrated lexicon (see Bijeljac-babic et al., 1997; Lemhöfer & Dijkstra, 2004; van Hell & Dijkstra, 2002; van Heuven et al., 1998; Voga & Grainger, 2007 for evidence supporting this claim). The model proposes that the bilingual mental lexicon comprises orthographic, phonological and semantic levels of representation. Regarding the representation and processing of high frequency and low frequency words, the BIA+ model assumes that orthographic/phonological word forms are stored with a resting activation level, which can be higher or lower depending on the frequency with which words have been processed. This resting activation level determines how fast words become active and how fast they are recognized and processed. High frequency words would have a higher resting activation level than low frequency words, and this would explain why the former are recognized and processed faster than the latter (see

Chapter 3 for more details). Turning to the representation and processing of cognate and non-cognate words, the model proposes that due to cognates' formal similarity across languages, their orthographic, phonological and semantic representations will be more strongly activated than those of non-cognates during lexical access. This overall stronger activation would cause cognates to be accessed and processed faster than non-cognates (see Chapter 4 for more details). The BIA+ model assumes that the word processing system interacts with the sentence parser, but it does not specify how exactly this might take place. Defining this is an important first step towards understanding how lexical and syntactic processing might interact during L2 syntax acquisition.

A hypothesis about how lexical processing (including processing of words differing in frequency and cognateness) could influence L2 syntactic processing is the Lexical Bottleneck Hypothesis (Hopp, 2018). This hypothesis proposes that a costly (e.g. slower) lexical processing may exhaust all the resources necessary to perform a native-like syntactic computation (see Chapter 3 and Chapter 4 for a full account)[5]. Cognates and high frequency words are processed faster than non-cognates and low frequency words, so this hypothesis predicts that a native-like syntactic processing should be more easily achieved when a structure includes the first pair of words than the second. The Lexical Bottleneck Hypothesis does not address, however, how words are integrated into the structure being processed, how sentences are processed in real time or how the influence of lexical processing on syntactic processing would hold for structures differing in cross-linguistic similarity. The studies supporting the hypothesis mostly involve processing of sentences that are temporarily ambiguous between an L1 structure and a different L2 structure (e.g. Hopp, 2017; Soares et al., 2018, 2019). In addition, being a hypothesis about processing, it does not address the influence of lexical processing on L2 syntax acquisition, although Hopp (2018) mentions that the hypothesis could potentially be applied to L2 acquisition.

In sum, investigating specific aspects of L2 processing and acquisition has allowed researchers to develop diverse hypotheses and theories and to conduct a large number of studies that have helped advance knowledge in particular research areas. However, investigating whether and how cross-linguistic syntactic similarity and lexical processing may facilitate initial L2 syntax acquisition, as is the goal of this dissertation, requires a framework that takes the highly specific research in different fields and integrates it into a single account of L2 acquisition. While the framework must be flexible enough to accommodate different theoretical proposals, it also needs to have some limits. That is, it must have clear principles that must be obeyed by any account that it incorporates (e.g. if the framework assumes that some of the properties of the human mind are innate, then it would not accommodate a theory that denies this idea). As advanced, I propose that the MOGUL framework is detailed and comprehensive

[5] Hopp does not specify which particular resources are responsible for difficulties in syntax processing. The hypothesis is presented as an extension to limited capacity models, which consider that "it is specifically L2 memory capacity, L2 decoding and/or L2 processing speed that are thought to underlie L2 grammatical difficulties" (McDonald, 2006, p. 383). It could be that these are (some of) the resources that the Lexical Bottleneck Hypothesis refers to.

enough to account for (i) the influence of L1 syntax on L2 syntax processing and acquisition, (ii) L2 word processing (including high/low frequency words and cognates/non-cognates) and (iii) the influence of this lexical processing on the acquisition of cross-linguistically similar and dissimilar L2 structures. In the next section, I give an overview of this framework and I further discuss these ideas.

3. The MOGUL framework

MOGUL is a theoretical framework dealing with how languages in the bilingual or multilingual mind are represented, processed and acquired in real time as well as with the interplay between languages and other cognitive systems (Sharwood Smith & Truscott, 2014, sec. 1.2). One of the main claims of the framework is that languages expand or diminish because of processing: a language develops when it is used, and linguistic ability declines in the absence of use. This idea is summarised in the *acquisition by processing principle,* which states that "acquisition is the lingering effects of processing" (Sharwood Smith & Truscott, 2014, sec. 4.2). MOGUL thus views (first and second) language acquisition from a processing-based perspective. Importantly, when Sharwood Smith and Truscott discuss how a *second* language is processed and acquired, they intend this discussion to apply to *second or other* languages*,* irrespective of the context in which they are learnt (Sharwood Smith & Truscott, 2014, sec. 1.3). In addition, the authors do not make a distinction between *acquisition* and *learning* as proposed by Krashen and, in fact, they use the two terms interchangeably (Sharwood Smith & Truscott, 2014, sec. 7.2.10). Finally, MOGUL is a model of how the *mind* (not to the brain) works. Therefore, the architecture of language and the specifics of language processing proposed do not purport to be equivalent to the representation and processing of language in the brain. Sharwood Smith and Truscott mention that harmonising psycholinguistic and neurolinguistic descriptions of language in the bilingual is a desirable but complex matter that is beyond the scope of their framework (Sharwood Smith & Truscott, 2014, sec. 3.4.5, 11.7). In what follows, I introduce MOGUL's architecture and I discuss how processing and acquisition by processing generally work in this framework.

3.1. Architecture

MOGUL takes a modular approach to the representation of language in the mind. The mind has long been proposed to be composed of processors or modules, each dedicated to a specific function, such as sight, hearing or memory (e.g. Fodor, 1983, 2000; Jackendoff, 1987, 1997; Pinker, 1994, 1997). In linguistics, Chomsky (1972, 1980) proposed the existence of a *language module* responsible for humans' linguistic ability. This language module was considered akin to body organs like the kidneys or the liver, in the sense that it was genetically specified, it served a specific function and it developed or grew in a natural way. Since Chomsky first advanced the idea, several authors have described language in the mind from a modular perspective (e.g. Fodor, 1983; Jackendoff, 1987, 1997, 1999, 2002; B. D. Schwartz, 1999; Smith & Tsimpli, 1995). Sharwood Smith and Truscott take Jackendoff's approach as the basis of MOGUL's architecture, even if their proposal differs from Jackendoff's in subtle ways (see Jackendoff, 2002 and Sharwood Smith & Truscott, 2014, sec. 2.3.2, 2.8.2 for more details).

MOGUL conceptualises the linguistic system as comprising three different subsystems or modules: phonological, (morpho)syntactic and conceptual. These modules are innate, in the sense that they have an invariant architecture that is biologically determined and can be found in all humans. The phonological module and the syntactic module, with connections to each other, form the so-called *core language system*. The conceptual module closely interacts with the core language system and is part of the *language broadly defined*. The core language system is domain-specific, for it is only associated with linguistic cognition (and not with other cognitive systems, e.g. perceptual systems). The modules that constitute it function using principles exclusive to phonology and syntax and employ their own linguistic code. Due to this domain specificity, the phonological and syntactic modules are more generally referred to as *the language module* (Sharwood Smith & Truscott, 2014, sec. 1.7.1). In addition, the core language system is species-specific, in the sense that it is exclusive to human beings. Sharwood Smith and Truscott acknowledge that other species, such as birds and apes, have forms of communication that could be seen as language, but maintain that these species cannot be claimed to have a language module like that of humans, biologically predetermined for language acquisition (Sharwood Smith & Truscott, 2014, sec. 11.4)[6]. On the other hand, the conceptual module is not exclusively linguistic and human. This module is responsible for encoding and interpreting the meanings of linguistic expressions —what is investigated by semantics and pragmatics— but it also attributes and decodes the meanings of visual images, smells and non-linguistic sounds. Consequently, the conceptual module is also involved in general cognition and, as such, associated to species other than humans. Finally, MOGUL describes three additional systems or modules that bear some relation to the language broadly defined: the auditory module, the visual module and the articulatory module. All these modules interact with the core language system during comprehension and production (see Figure 1.1). Yet, they also have non-linguistic functions, such as representing non-linguistic sounds and images (Sharwood Smith & Truscott, 2014, sec. 1.7.1). These non-linguistic sounds and images can be recognised and processed thanks to the existence of direct connections between the auditory and the conceptual modules and between the visual and the conceptual modules (Sharwood Smith & Truscott, 2014, sec. 1.7.1, 5.2.3).

[6] For example, some birds create and learn songs in a way that resembles human language acquisition (see Bolhuis et al., 2010; Prather et al., 2017 for reviews of behavioural and neurobiogical evidence). Nevertheless, Sharwood Smith and Truscott mention that birds only utilise what they perceive to develop their "language" and this does not seem to be the case for humans (*poverty of stimulus argument*, Chomsky, 1980; Hornstein & Lightfoot, 1981). In addition, mynah birds and parrots identify and imitate patterns of human speech (Klatt & Stefanski, 1974; Pepperberg, 1981). However, even if they are exposed to human language, this will never lead to language acquisition. Finally, some apes use vocalisations to escape from predators, defend against other animals or find food. Yet, there is no evidence that they can learn their vocalisations just as some birds learn their songs, and apes interacting with humans do not learn new ways of vocalising to communicate in human habitats (Tomasello, 2007).

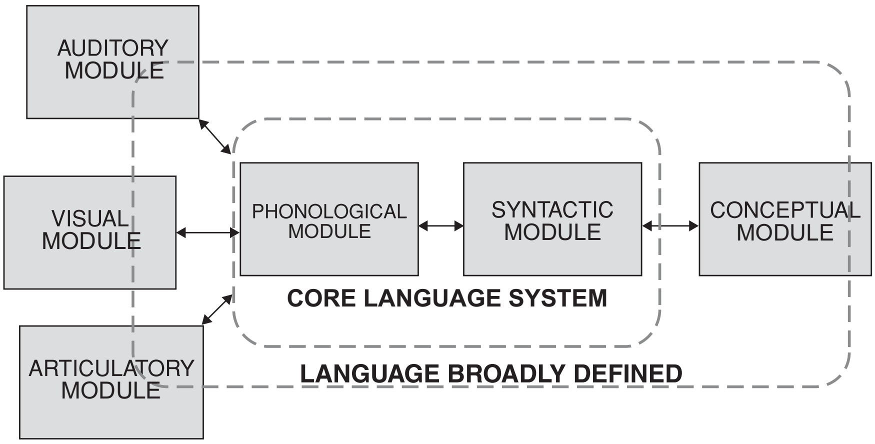


FIGURE 1.1. MOGUL's core language system (the phonological module and the syntactic module, connected to each other) and the language broadly defined (comprising most of the conceptual module and part of the auditory module, the visual module and the articulatory module). Adapted from Sharwood Smith and Truscott (2014, p. 17).

Each module in the linguistic system has two basic components: an *information* or *lexical store* and a *computational system* or *processor.* Thus, the phonological module contains a phonological store and a phonological processor, the syntactic module contains a syntactic store and a syntactic processor, and so on. Information stores are sublexicons containing all elements and combinations of elements constituting linguistic knowledge, which in MOGUL are called *structures* or *representations*[7]. To name a few, representations include individual phonemes and phonological word forms in the phonological store, lexical categories (e.g. [N], [V]), functional categories (e.g. Inflection, Case) and syntactic structures in the syntactic store and semantic features and conceptual roles (in generative terms, theta roles, e.g. agent, patient) in the conceptual store. Some of these representations, such as lexical and functional categories, are innate to the stores. Others, such as syntactic structures, are built by experience. The lexical stores are also the locus of the activity of the processors, so they are additionally considered memory stores or blackboards where representations are written. Specifically, the processors select, combine and integrate representations in the lexical stores into larger representations following the principles of each module (Sharwood Smith & Truscott, 2014, sec. 1.7.1, 2.3.2, 2.3.3)[8]. Each processor only recognizes and manipulates the representations in its module. To allow communication across modules, MOGUL proposes a set of *interfaces.* Interfaces connect representations in adjacent modules by means of indexes

[7] Sharwood Smith and Truscott use the labels AS, VS, PS, SS and CS to refer to acoustic, visual, phonological, syntactic and conceptual structures or representations. I will not use these labels and I will use the term *representation* instead of *structure* to avoid the confusion between grammatical or syntactic structures, which are the object of study of the dissertation, and other contents of the lexical stores.

[8] These principles are seen as genetically determined constraints on the way in which languages may develop, what in generative research has been referred to as Universal Grammar (Sharwood Smith & Truscott, 2014, sec. 1.7.1, 2.2.2).

to form *chains of representations*, with most of these chains being formed through experience[9]. Linguistic elements stored as chains of representations include words, derivational affixes, idioms, collocations, constructions, and more. For example, the chain of representations for the lexical item *bed,* which I have coindexed with the arbitrary index j, includes the phonological representation $/bed/_j$, syntactic representations indicating the word's lexical category (e.g. $[N_j]$) and number (e.g. $[Singular_j]$) and a conceptual representation along the lines of PIECE OF FURNITURE USED FOR SLEEPING ON_j (Figure 1.2). Additionally, this chain of representations is coindexed with the acoustic and orthographic representations of the word *bed* and with any sounds and images associated with that object (Sharwood Smith & Truscott, 2014, sec. 1.7.2).

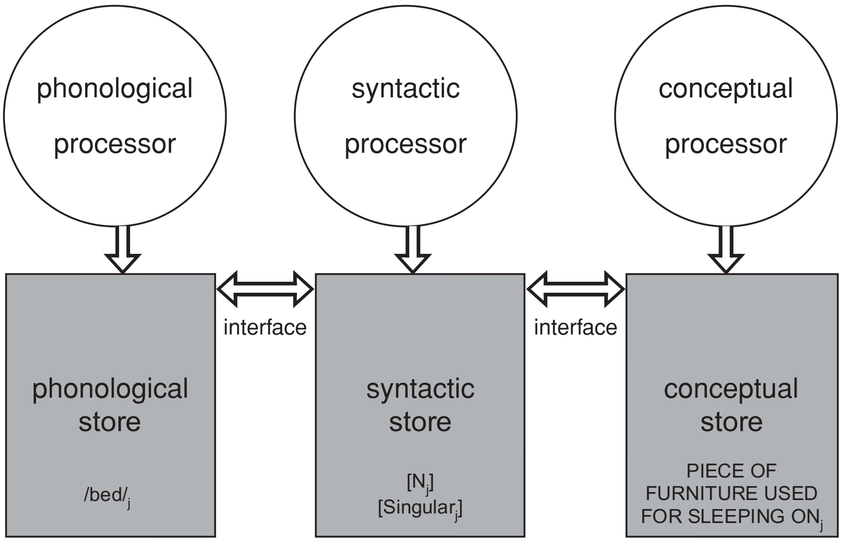


FIGURE 1.2. Structure of the phonological, syntactic and conceptual modules in the linguistic system. For simplicity's sake, other modules part of the language broadly defined are not included, but the same structure is assumed. The figure also illustrates the chain of phonological, syntactic and conceptual representations corresponding to the word *bed* in each module, connected with the index j. Adapted from Sharwood Smith and Truscott (2014, p. 36).

[9] These indexes are not as the ones used in linguistics to indicate coreference or the link between a trace and an element that has moved. The indexes used by interfaces have no function besides connecting representations across modules (Sharwood Smith & Truscott, 2014, sec. 2.3.3).

3.2. Language processing

As previously mentioned, MOGUL conceives language acquisition as the lingering effects of processing (*acquisition by processing principle*). To comprehend this principle and, ultimately, how L2 syntax acquisition works in MOGUL, it is convenient to understand first how language processing is conceptualized in this framework. MOGUL can account for language comprehension and production. In this section, I focus on language comprehension. The process underlying production is simply the reverse to the one underlying comprehension (see Sharwood Smith & Truscott, 2014, sec. 3.4.1 for more information).

When a speaker is exposed to spoken language (e.g. a word, a sentence), an acoustic representation of the input activates in the auditory module. The interface between the auditory and the phonological modules activates coindexed representations in the phonological store. Subsequently, the interface between the phonological and the syntactic modules activates coindexed representations in the syntactic store and, then, the interface between the syntactic and the conceptual modules activates coindexed representations in the conceptual store (Sharwood Smith & Truscott, 2014, sec. 3.4.1)[10]. MOGUL assumes that language processing is incremental, a generally accepted fact in the processing literature (Altmann & Steedman, 1988; Traxler & Pickering, 1996; van Gompel & Pickering, 2007; Williams, 2006). Specifically, when a representation in a given module activates, an interface immediately activates coindexed representations in an adjacent module, i.e. it does not wait for the activity in a module to finish before activating items in a contiguous one. Consequently, processing occurs in parallel in the phonological, syntactic and conceptual modules. This incremental activation of representations causes that the processors also work incrementally, trying to construct an appropriate representation of the input based on the items that become active in each store (Sharwood Smith & Truscott, 2014, sec. 3.3.4).

Importantly, during input processing not only the target phonological, syntactic and conceptual representations activate in their respective lexical stores, but representations sharing features with the target ones also activate and compete to be selected by the processor in each module. The representation selected by each processor is the suitable representation that is most active at each moment of the processing activity. More precisely, MOGUL proposes that all items in a lexical store have a *resting activation level* and a *current activation level.* On the one hand, an item's resting activation level is the result of its use in processing: the more frequently an item is used, the higher its resting activation level is. The resting level is the starting activation level for each representation and the higher the resting level is, the faster a representation becomes available for processing. On the other hand, the

[10] If a speaker is exposed to written language, an orthographic representation activates in the visual module. Then, an interface between the visual and the auditory modules activates coindexed acoustic representations (Sharwood Smith & Truscott, 2014, sec. 5.4.4).

current activation level of a representation is a combination of its resting level and the activation it receives during a processing event. The more features a representation shares with the input, the higher the activation it receives. The representation with the highest current activation level is typically the one that wins competition against any alternatives in a lexical store and is selected for processing (Sharwood Smith & Truscott, 2014, sec. 3.3.5, 3.4.1). The idea of competition and selection based on activation levels is common in the processing literature and has been adopted in many models of word and sentence processing, including comprehension and production (e.g. Dell, 1986; Dijkstra & van Heuven, 1998, 2002; MacWhinney, 1987a; McClelland & Rumelhart, 1981; Morton, 1969)[11].

Additionally, processing is not simply a matter of interfaces activating phonological representations, then syntactic representations and finally conceptual ones. As shown in Figure 1.2, interfaces are bidirectional, so they continually activate representations in the modules on each of their sides. For instance, as input is incrementally processed, active representations in the conceptual module raise the current activation level of coindexed syntactic representations, which may or may not coincide with the ones initially activated by the interface between the phonological and the syntactic modules and which may influence the syntactic analysis being conducted. Therefore, syntactic and semantic processing occur separately, but do interact (Sharwood Smith & Truscott, 2014, sec. 3.4.1, 3.4.4). This aligns with constraint-based lexicalist models of processing, which assume that syntactic and semantic information are concurrently used for processing, influencing one another to converge on the most suitable parse (e.g. MacDonald et al., 1994; McClelland et al., 1989; Trueswell & Tanenhaus, 1994). By contrast, this conception of syntactic and semantic processing goes against garden-path or syntax-first models, which assume that the syntactic processor produces a single analysis first, and that semantics is consulted later, causing the first analysis to be either accepted or abandoned (e.g. Frazier, 1987, 1989; Frazier & Clifton, 1997; Frazier & Fodor, 1978; Frazier & Rayner, 1982).

Finally, a word must be said about the result of processing. Once processing has terminated, the current activation level of the chain of representations ultimately used to process the input (e.g. the word, the sentence) falls back towards the resting activation level, landing slightly above where it was prior to the processing event. In other words, each time a representation receives stimulation, its resting activation level rises. Representations having a high current activation level will land at a higher resting activation level than representations having a lower current activation level. Put differently, the increase in resting activation level derived from processing depends on the strength of the stimulation experienced (Sharwood Smith & Truscott, 2014, sec. 3.3.5). This has consequences for future processing, since, as advanced, the higher the resting activation level of a representation is, the more readily

[11] Sharwood Smith and Truscott acknowledge the need to understand how the concepts of activation and activation levels, as described in MOGUL, could be mapped onto neural activation during speech processing. However, as mentioned, their framework is not intended to describe neural architecture (Sharwood Smith & Truscott, 2014, sec. 3.4.5).

available it is for processing. This idea will also be crucial for acquisition by processing, as will be discussed in the next section.

3.3. Acquisition by processing

MOGUL's approach to language acquisition, summarised in the *acquisition by processing principle*, is framed within what Sharwood Smith and Truscott label *Acquisition by Processing Theory* (APT) (Sharwood Smith & Truscott, 2014, sec. 4.2). The authors are not the first to suggest that acquisition should be linked to processing. The APT is inspired by the work of Carroll (1999, 2001), who, as mentioned in Section 2, proposed that acquisition was the result of processing failure. Specifically, she claimed that when processing mechanisms could not parse the input, acquisition mechanisms triggered learning. Sharwood Smith and Truscott do not agree with the distinction between success and failure in processing and point out that, in those cases in which the input is difficult to process, parsing is not abandoned. Instead, the processing system usually makes some adjustments and derives a message from the input (Sharwood Smith & Truscott, 2014, sec. 4.2.1). Accordingly, they propose that acquisition should be understood in terms of processing that does something new and produces long-lasting changes in the linguistic system. The term *acquisition* connects with an idea of language development that involves incorporating some entities (e.g. words, structures) from the input into the linguistic system. Although the authors agree on using this word due to its widespread use in the literature, they prefer to think about language as simply developing or growing after processing (Sharwood Smith & Truscott, 2014, sec. 4.2).

According to the APT, upon exposure to an element that has never been encountered before, the phonological, syntactic and conceptual processors establish a new representation for it in the phonological, syntactic and conceptual stores, forming a chain of representations. This is done for the sole purpose of processing the input and, as such, is a processing mechanism, not a learning mechanism. For instance, when a speaker is exposed to a syntactic structure for the first time, the input from the phonological module incrementally activates coindexed items in the syntactic store. The syntactic processor selects and combines the most active items into an appropriate representation following syntactic principles, just as it would do to process the structure if it were already part of the linguistic system. This novel syntactic representation is coindexed with the phonological and conceptual representations of the words that constitute it as encountered in the input. Once a representation is created as a result of processing, it is stored in the linguistic system, initially with a very low resting activation level. If the representation is not used again in processing (in the previous example, if a sentence with that syntactic structure is not used in comprehension or production), its resting activation level will fade away and the representation will not become a permanent part of the lexical store. In familiar thinking terms, it will not be acquired. By contrast, every time the representation is used in processing, its resting activation level will increase (Figure 1.3). As described in the previous section, after each processing event, a representation's current activation level gradually declines towards its resting activation level, stopping at a slightly higher position than the original one. Crucially, the higher a representation's resting activation level is, the more firmly established it is in a lexical store. Thus, if a representation is repeatedly used in processing, it will gradually become a more stable item in the linguistic system, potentially remaining there indefinitely. In familiar thinking terms, it will be acquired (Sharwood Smith & Truscott, 2014, sec. 4.2).

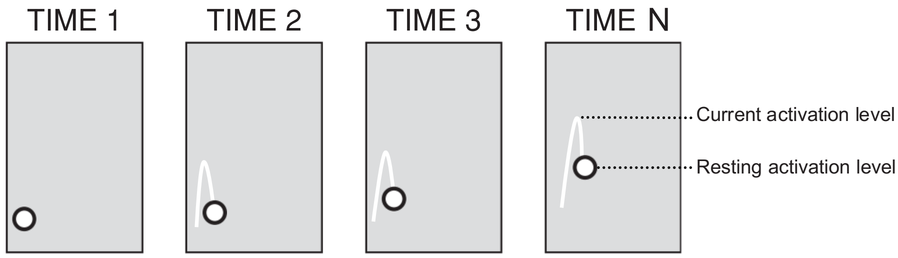


FIGURE 1.3. Acquisition by processing in MOGUL: Successive processing of an item involves raising its resting activation level on each occasion, which causes it to become firmly established in the linguistic system. Adapted from Sharwood Smith and Truscott (2014, p. 95).

3.4. The L2 in the bilingual mind

So far, I have not commented on whether and how the representation, processing and acquisition of language described would vary between the L1 and the L2. This is because, in the MOGUL framework, first and second languages are similar organisms, represented in the same architecture, and processed and acquired in a similar way. Nevertheless, Sharwood Smith and Truscott acknowledge that there are differences between L1 and L2 acquisition and attribute this to the fact that the L1 and the L2 coexist in the bilingual mind and compete to be processed by a single processing system (Sharwood Smith & Truscott, 2014, sec. 3.5). This goes in line with evidence that the same brain networks are involved in processing the L1 and the L2 (see Del Maschio & Abutalebi, 2019 for a review) and with behavioural and neurophysiological evidence that all the languages that a speaker knows activate during processing (e.g. lexical co-activation: Blumenfeld & Marian, 2013; Dijkstra et al., 1999; Marian & Spivey, 2003; van Heuven et al., 1998; syntactic co-activation: Luque et al., 2018; Sanoudaki & Thierry, 2015; Thierry & Sanoudaki, 2012; Vaughan-Evans et al., 2020). This conception of the bilingual mind raises several questions, such as how the two languages of a bilingual are differentiated and how the L1 influences L2 acquisition and processing, including initial L2 syntax acquisition, the focus of this dissertation. In what follows, I address the first question and postpone the discussion of the second question until the next section.

The bilingual mind must have some way of marking the language to which words, structures and other linguistic items belong. In MOGUL, this could be done in two different ways. The first option is that acoustic/orthographic, phonological and syntactic representations carry a *language tag*, i.e. an index identifying the linguistic identity of each representation (*language-tagging hypothesis,* Sharwood Smith & Truscott, 2014, sec. 6.4.2). Additionally, conceptual representations specifically linked to the culture of a language community would also bear a language tag (e.g., the conceptual representation of *cricket* would be tagged as English). The language-tagging option could only be accepted if language tags were seen as a generic form of connecting representations across modules. In this case, upon exposure to spoken or written input in one of the languages of the bilingual, appropriate acoustic and/or orthographic representations with the corresponding language indices would activate. This would trigger the activation of phonological, syntactic (and conceptual) representations tagged with that very same language index. By this process, the representations of the target language would have higher current activation levels than the representations of the non-target language and, thus, they would be most likely to be selected for processing. The existence of some form of language tags is present, for instance, in models of bilingual word comprehension and acquisition (e.g. Dijkstra & van Heuven, 2002; Grainger et al., 2010) and in accounts of the representation and development of syntactic structures in bilinguals (e.g. Hartsuiker et al., 2004; Hartsuiker & Bernolet, 2017; Hartsuiker & Pickering, 2008).

As an alternative to the language-tagging hypothesis, Sharwood Smith and Truscott propose the *conceptual triggering hypothesis*. This hypothesis postulates that the phonological and syntactic modules are blind to the language that they are processing (see Dekydtspotter, 2001; MacSwan, 2000 for similar claims) and that language differentiation occurs outside the core language system, in the conceptual module (Sharwood Smith & Truscott, 2014, sec. 6.4.3). Specifically, the conceptual store would contain a conceptual representation of each language that a speaker has some knowledge of (e.g. representations such as SPANISH and ENGLISH for Spanish-English bilinguals). These conceptual representations would be directly associated with particular sounds and prosodic patterns in the auditory module and with particular orthographic representations in the visual module, by means of direct interfaces between the auditory/visual and conceptual modules (i.e. bypassing the core language system). Consequently, there would be no language tags in the phonological or the syntactic modules marking representations as belonging to a language or another. The association of phonological and syntactic representations with a specific language would be the result of their connection to acoustic or visual representations coindexed with a language concept. In this case, each time that an acoustic or an orthographic representation activates, the conceptual representation of its associated language would activate as well. Then, the acoustic or orthographic representations would activate phonological and syntactic representations coindexed with them. Overall, the conceptual representation of the target language, together with the language-specific sounds and orthographic features, would selectively activate the representations of this language more strongly than any representations of the non-target language.

3.5. The influence of L1 syntax on initial L2 syntax acquisition

The L1 influences L2 acquisition in both positive and negative ways (e.g. Chang & Zheng, 2015; Henry et al., 2009; McManus, 2015; Tokowicz & MacWhinney, 2005). As advanced, one of the main aspects that I study in this dissertation is whether L1 syntax can facilitate adults' initial acquisition of L1-L2 similar structures, compared to dissimilar structures. Traditionally, the term used to refer to the influence of the native language on second language acquisition is language *transfer.* This term was introduced in the mid-twentieth century by the behaviourist theory, which conceived language learning as the development of particular habits through practice and reinforcement (Skinner, 1957). When it comes to second language learning, behaviourists such as Lado (1957) assumed that learners carried over habits from their L1, which could facilitate learning —if the habit to be learnt was similar to the inherited one— or hamper it —if the L2 habit and the L1's differed. These two processes were named *positive transfer* and *negative transfer*, respectively. The word *transfer* continues to be used in the 21st century (e.g. Bokander, 2020; Guo, 2022; Hartsuiker & Bernolet, 2017; Hopp et al., 2019; Tolentino & Tokowicz, 2014; Westergaard, 2021; Yu & Odlin, 2016). However, despite its widespread use, *transfer* has been argued not to be the most adequate term to refer to the influence of the native language on second language development (see Chapter 2). In MOGUL, Sharwood Smith and Truscott prefer to use the term *cross-linguistic influence* (Kellerman & Sharwood Smith, 1986; Sharwood Smith, 1983).

To investigate whether and how L1 structures influence initial learning of L2 structures, it is necessary to define first the initial state of L2 acquisition. This was extensively studied in the 1990s, especially within theories that assumed that language acquisition was guided by Universal Grammar (UG)[12]. A debate was held about whether the L2 learner started the acquisition process just with UG, as in first language acquisition (Epstein et al., 1996, 1998; Platzack, 1996) or with the L1 grammar. The accounts supporting the second view differed in the grammatical representations they assumed to be available at the initial state of L2 acquisition. Some assumed that the L2 learner had access to the whole of L1 grammar (the FT/FA model, B. D. Schwartz & Sprouse, 1994, 1996). Specifically, a copy of the L1 grammar was made at the start of L2 acquisition and, whenever necessary, it was modified to meet the characteristics of L2 grammar[13]. Others proposed that the initial state of L2 acquisition consisted of a grammar with L1 lexical categories but without functional categories, which developed gradually as a result of L2 processing (*Minimal Trees Hypothesis*, Vainikka & Young-Scholten, 1994, 1996a, 1996b). A third proposal was that L2 acquisition started with the lexical and functional categories of the L1, but that feature values were not specified and had to be acquired from L2 input (*Valueless Features Hypothesis*, Eubank, 1993, 1994, 1996). In all cases, UG constrained the development of L2 grammar.

MOGUL also assumes that UG guides language acquisition. UG is not conceived as a separate entity in the linguistic system, but as a genetic base specifying the initial state of the

[12] Universal Grammar is the biological component of the Faculty of Language, which allows acquiring and using particular languages (Berwick & Chomsky, 2016; Chomsky, 1995).

[13] B.D. Schwartz and Sprouse did not specifically use the term *copying* to refer to the process by which L1 grammar becomes the initial state of L2 grammar. White (2003) named the process *copying* and later B.D. Schwartz and Sprouse (2021) acknowledged it as correct.

phonological and syntactic stores, the interfaces connecting the phonological and syntactic modules between them and with other modules, and the principles by which the phonological and syntactic processors construct representations (Sharwood Smith & Truscott, 2014, sec. 2.3.1, 2.2.2). Regarding the initial state of L2 acquisition, MOGUL is most sympathetic with the FT/FA model. Specifically, Sharwood Smith and Truscott propose that the initial state of L2 acquisition consists of the processors and information stores innate in the linguistic system and that, when L2 acquisition starts, the information stores contain all L1 representations, with a high resting activation level due to their widespread use in the L1 (Sharwood Smith & Truscott, 2014, sec. 10.3).

In the previous section, I advanced that linguistic elements from all the languages that a speaker knows are active and can be used during processing, irrespective of the language of the input. This implies that, when encountering the L2 for the first time, the processors will attempt to process the input using compatible L1 representations in the lexical stores. Following the Acquisition by Processing Theory, in those cases in which processing requires a processor to construct a new representation (e.g. if an appropriate representation of the input is not present in the linguistic system), this will be done, simply as a means of processing the language (Sharwood Smith & Truscott, 2014, sec. 4.2). Considering all this, I deduce that, if an L2 structure can be parsed using an L1 syntactic representation, as in the case of cross-linguistically similar structures, this will be done, and no new representation will be created. This L1 syntactic representation will neither be "transferred" nor "copied" to the L2; the syntactic processor will have direct access to the representation in the syntactic module, shared for the L1 and the L2. By contrast, if no L1 representation is adequate to process the input, as in the case of cross-linguistically dissimilar structures, a suitable syntactic representation will need to be created. Simply put, cross-linguistically similar structures will be part of the linguistic system from the beginning of L2 acquisition, but cross-linguistically dissimilar structures will have to be learnt from input[14]. This view is generally compatible with the FT/FA model and with other theories and models of L2 acquisition, as mentioned in Section 2. I will develop these ideas in more detail in Chapter 2.

3.6. The influence of lexical processing on initial L2 syntax acquisition

Apart from the potentially beneficial role of cross-linguistic syntactic similarity in the initial acquisition of L2 structures, the second main aspect I investigate in this thesis is whether and how lexical processing facilitates the initial acquisition of L1-L2 similar and dissimilar structures. Specifically, I explore the effect of two pairs of words that differ in processing speed (and accuracy), namely high frequency vs. low frequency words and cognates vs. non-cognates. I previously mentioned that accounts of word processing such as the BIA+ model attribute this processing advantage to a difference in activation between high and low

---

[14] It is common practice to use the term *acquisition* for cross-linguistically similar structures and morphosyntactic features even if, from a theoretical perspective, these are assumed to be present at the beginning of L2 acquisition and, thus, strictly speaking do not need to be acquired. I will also follow this practice in my dissertation.

frequency words, and between cognates and non-cognates. I argue that an explanation based on activation levels can also be accommodated within the MOGUL framework.

On the one hand, Sharwood Smith and Truscott explicitly mention how frequency of occurrence is reflected in MOGUL. Like in the BIA+ model, each time that an item (e.g. a word) is processed, its resting activation level increases (cf. Figure 1.3). Following lexical access research (e.g. B. Gordon & Caramazza, 1985), Sharwood Smith and Truscott assume that the increase in resting activation level after each use is large when an item is at the first stages of development, but gradually reduces until, at some point, the resting activation level increases no more. This relation between frequency of occurrence and resting activation level applies to all types of representations in the lexical stores, not only to lexical items (Sharwood Smith & Truscott, 2014, sec. 4.6.5). In any case, since high frequency words are processed more often than low frequency words, the first would have a higher resting activation level than the second. As mentioned in Section 3.2, the resting activation level establishes how quickly representations become available for processing (Sharwood Smith & Truscott, 2014, sec. 3.3.5). Consequently, the higher resting activation level of high frequency words compared to low frequency words would cause the former to be processed faster than the latter. Due to their higher resting activation level, high frequency words would also have a higher current activation level than low frequency words. This is because, as advanced, an item's current activation level is its resting level plus the activation received during the ongoing processing. The items with the highest current activation level are most likely to be included in the representation of input under construction (Sharwood Smith & Truscott, 2014, sec. 3.3.5). Thus, the higher current activation level of high frequency words compared to low frequency words would cause that the former are selected by the processing system more often than the latter and, hence, that they are processed more accurately. Since the frequency effects described are the result of words being processed more or less often, the question arises regarding how lexical frequency would be relevant during the initial stage of L2 acquisition, i.e. when learners have never been exposed to the L2 and, hence, have never processed L2 words with different frequencies. I propose that lexical frequency could influence processing at the earliest stage of L2 acquisition if, for instance, the lexical items differing in frequency were cognates in the L1 and in the L2 and the frequency that varied was that of the lexical items in the L1. In this case, the L2 cognate words would be processed as if they were equivalent to their L1 counterparts, with a higher or lower frequency of occurrence. These ideas will be discussed in more detail in Chapter 3.

On the other hand, Sharwood Smith and Truscott do not discuss how cognates would be represented and processed in MOGUL. Nevertheless, I argue that the way in which the framework describes the representation and processing of lexical items could accommodate an account of cognate representation and processing. In brief, MOGUL assumes that during processing, all representations sharing features with the input activate, irrespective of the language to which they belong (Sharwood Smith & Truscott, 2014, sec. 3.3.5, 3.4.1). This implies that, since cognates share orthographic and/or phonological features across two or more languages, when a bilingual processes a cognate, the orthographic and/or phonological representation of the word in two languages would activate. By contrast, when the bilingual processes a non-cognate, the orthographic and/or phonological representation of the non-cognate's equivalent in the non-target language would not activate, for the two lexical items share no formal features. In both cases, activation would spread from the orthographic/phonological representations to coindexed syntactic and conceptual representations, the latter being shared for translation equivalents. Crucially, activation of this conceptual representation would be stronger for cognates than for non-cognates for, in the first case, it would receive activation from two chains of orthographic, phonological and syntactic representations. Consequently, a stronger activation would spread back to cognates' than to non-cognates' syntactic, phonological and orthographic representations. In other words, the chain of representations of cognates would have a higher current activation level than that of non-cognates. Additionally, cognates would have a higher resting activation level than non-cognates since, as mentioned in Section 3.2, the higher the current activation level of a representation is, the higher the resting activation level derived from processing (Sharwood Smith & Truscott, 2014, sec. 3.3.5). The higher current and resting activation level of cognates compared to non-cognates would explain why the former are processed faster and more accurately than the latter. The difference in cognate vs. non-cognate processing should also hold during the early stages of L2 acquisition, once learners have learnt the cognate and non-cognate vocabulary. These ideas will be developed further in Chapter 4.

Importantly, in MOGUL the activation of words directly influences the processing (and by extension, the acquisition by processing) of syntactic structures. As the words in a sentence are incrementally processed, the syntactic representations of these words activate in the syntactic store. Activation then spreads from the words' representations to larger syntactic representations containing them, raising their current activation levels. For example, when a speaker encounters the sentence *Kick the ball,* processing of *kick* involves activating its syntactic representation (e.g. $[V_i]$), which subsequently activates other syntactic representations with $[V_i]$, possibly including $[_{CP}\ [_{VP}\ V_i\ NP]]$ (Sharwood Smith & Truscott, 2014, sec. 4.5.3)[15]. For novice L2 learners, this syntactic representation may correspond to a structure acquired as part of the L1 or to an L2 structure just established in the linguistic system, having a low resting activation level. Anyhow, MOGUL assumes that when activation spreads within a lexical store (in this case, within the syntactic store), the current activation level of the representation from which activation spreads (e.g., $[V_i]$) influences the extent of the rise in the current activation level of representations containing it (e.g., $[_{CP}\ [_{VP}\ V_i\ NP]]$) (Sharwood Smith & Truscott, 2014, sec. 3.4.1). This implies that the higher the current activation of a representation is, the larger the increase in the current activation of the

[15] The subindex i is arbitrary and identifies the chain of representations corresponding to the word *kick* (cf. Section 3.1). In addition, Sharwood Smith and Truscott choose to use the label NP instead of DP, which is the preferred category in some linguistic theories, to ease comprehension by all readers regardless of their familiarity with these theories (2014, p. 45).

representations including it. If high frequency words and cognates have a higher current activation level than low frequency words and non-cognates, then processing a structure with the first two types of words should result in a higher activation of the structure than processing the same structure with the last two types of words.

When processing terminates, the current activation level of the structure decays towards the resting activation level, landing at a position slightly above the original (Sharwood Smith & Truscott, 2014, sec. 3.3.5). As mentioned, the increase in a representation's resting activation level depends on the strength of its current activation level, so that a representation that has a high current activation level will end up having a higher resting activation level than a representation that has a lower current activation level. Consequently, I propose that a structure processed with a high frequency word or a cognate will have a higher resting activation level than the same structure processed with a low frequency word or a non-cognate. This has consequences for acquisition by processing, since the higher an item's resting activation level is, the more firmly it is established in the linguistic system (Sharwood Smith & Truscott, 2014, sec. 4.2). In Chapters 3 and 4, I will further develop how lexical frequency and cognateness could affect L2 syntax acquisition and whether and how this could be modulated by the cross-linguistic similarity of structures.

4. Aims and outline of the dissertation

In this section, I summarise the aims of the dissertation and I briefly outline the structure of the following chapters. The main goal of this dissertation is to study two factors that could influence initial L2 syntax acquisition to determine their potentially facilitative role. Specifically, I investigate (a) whether and how cross-linguistic similarity between L1 and L2 structures facilitates their initial acquisition (Chapter 2) and (b) whether and how lexical processing facilitates the acquisition of cross-linguistically similar and dissimilar L2 structures by comparing the effect of (i) high vs. low frequency words (Chapter 3) and (ii) cognate vs. non-cognate words (Chapter 4). Research on the facilitative role of cross-linguistic syntactic similarity at the earliest stage of adult L2 syntax acquisition is virtually absent (but see Tolentino & Tokowicz, 2014) and, to my knowledge, the facilitative role of lexical frequency and cognates has not been studied. I expect this dissertation to shed some light on these topics by means of five behavioural experiments conducted with complete novice L2 learners. Several theories of second language acquisition directly or indirectly address the influence of L1 syntax on the acquisition of L1-L2 similar and dissimilar structures. By contrast, as far as I know, none specifies how processing cross-linguistically similar and/or dissimilar structures with words differing in lexical frequency or cognateness may affect their acquisition. By elaborating on how a theoretical framework like MOGUL could generate and test hypotheses about these aspects, I additionally expect this dissertation to contribute towards overcoming this limitation.

- In **Chapter 2**, I examine whether Spanish natives without knowledge of Galician demonstrate a learning advantage for a cross-linguistically similar structure (an embedded clause existing in Spanish and Galician) over a cross-linguistically dissimilar one (an embedded clause expressing the same meaning, but formally different from its Spanish counterpart). In Experiment 1, I used an implicit learning paradigm. Participants were aurally exposed to the two structures in sentences made up of Spanish-Galician cognates while conducting a Plausibility Judgement Task (PJT). Learning was tested in an auditory Grammaticality Judgement Task (GJT). Results showed that learners had some sensitivity to the difference between sentences formed by the cross-linguistically similar structure and sentences violating it, which is taken as evidence that the similar structure was established in learners' linguistic system. By contrast, learners showed no sensitivity to the difference between sentences formed by the cross-linguistically dissimilar structure and sentences violating it, providing no evidence that the dissimilar structure had been learnt. In an attempt to facilitate learning, in Experiment 2 I turned to an explicit learning paradigm. Participants were exposed to the same structures as in Experiment 1 but the length of the exposure phase increased and, instead of a PJT, they conducted an auditory and visual structure-search task. Learning was tested in an auditory and visual GJT with feedback. In this case, the test provided evidence that the cross-linguistically similar and dissimilar structures were both part of learners' linguistic system. As predicted, learning was greater for the similar structure than for the dissimilar structure.

- In **Chapter 3,** I focus on whether and how learning of the Spanish-Galician similar and dissimilar structures in Chapter 2 is affected by differences in lexical frequency. In Experiment 3, I replicated Experiment 2 but I varied the lexical frequency of the Spanish equivalents of the cognate verbs in the embedded clauses, so that, overall, the verbs in Experiment 3 had a significantly lower frequency than those in Experiment 2. Then, I compared the results of Experiments 2 and 3. Participants learning the structures with high frequency verbs (Experiment 2) were significantly more sensitive to the cross-linguistically dissimilar structure than participants learning the structures with low frequency verbs (Experiment 3). This suggests that high frequency verbs facilitated the acquisition of the dissimilar structure. By contrast, sensitivity to the cross-linguistically similar structure did not differ between experiments, suggesting that lexical frequency did not facilitate its learning. The learning advantage for the similar structure over the dissimilar one obtained with high frequency verbs (Experiment 2) was also found with low frequency verbs (Experiment 3). However, the advantage was larger for structures learnt with low frequency verbs than with high frequency verbs. That is, learning the structures with high frequency verbs, as opposed to low frequency verbs, reduced the learning distance between the similar and the dissimilar structures.

- In **Chapter 4,** I study the effect of cognateness in the initial acquisition of cross-linguistically dissimilar structures (Experiment 4) and similar structures (Experiment 5).

In Experiment 4, I used an implicit learning paradigm to compare how two groups of Spanish natives with no knowledge of Basque learnt a mini-language with Basque vocabulary (Spanish-Basque non-cognate nouns and either cognate or non-cognate verbs) and two Basque-based structures (having SOV or OSV word order and postpositional agent-patient marking). The two groups of participants first learnt the nouns and the cognate or non-cognate verbs with pictures. Then, they were aurally and visually exposed to the structures via sentence-picture pairs. One of the groups was exposed to sentences with cognate verbs and, the other, to sentences with non-cognate verbs. Learning was tested in a sentence-picture congruency task and in a written picture-description task. Exposure to the structures with cognates, as opposed to non-cognates, resulted in a more accurate use of these structures in the picture-description task. This suggests that cognates facilitated learning of the cross-linguistically dissimilar structures. In Experiment 5, I investigated whether this facilitation varied when the structures of the mini-language were similar to Spanish grammar. Specifically, the structures studied had SVO or OVS word order and prepositional patient marking. In this case, being exposed to the structures in sentences with cognates, as opposed to non-cognates, did not lead to a significantly more accurate performance, nor in the congruency task nor in the picture-description task. In other words, the facilitative role of cognates in Experiment 4 did not replicate for cross-linguistically similar structures in Experiment 5.

- In **Chapter 5**, I summarise the findings and contributions of this dissertation and I present some ideas for future research.

Chapter 2

The facilitative role of cross-linguistic syntactic similarity in initial L2 syntax acquisition

1. Introduction

In second language acquisition research, the initial state or the linguistic knowledge that the second language learner starts with has been an important topic of research. In the last decades, researchers have come to agree on the fact that learners have access to the native language and that the L1 plays a role in second language development, which can be positive or negative (Alonso, 2016; Bardovi-Harlig & Sprouse, 2017; Jarvis & Pavlenko, 2008; MacWhinney, 2005; McManus, 2015, 2022; Ringbom, 2007, among others). As mentioned in Chapter 1 (Section 3.5), the term that has been traditionally employed to refer to the influence of the L1 on L2 acquisition is *transfer.* In the conventional sense, *transfer* evokes a change of location. When applied to second language acquisition, the notion of *transfer* suggests that the element of the native language being transferred (a category, a feature, a structure, or other) is removed from the native language and implemented somewhere else, depriving the former of this element. This is a deceptive implication; at most, the term *transfer* describes a process of *copying* or *cloning* by which an element from the first language becomes present in the second language, without impoverishing the former in any way (Sharwood Smith & Truscott, 2006). However, the copying metaphor is not without its problems. As Sharwood Smith and Truscott point out, a central issue is defining what triggers the copying. It could be assumed that copying occurs when a learner is exposed to unfamiliar input. Yet, this poses the question of what *unfamiliar* means*:* A new dialect? A new accent? A new language? As an alternative to the problematic notions of *transferring* and *copying*, Sharwood Smith (1983) suggested the term *cross-linguistic influence,* which nowadays is very often used in second language acquisition research*.*

Researchers in the field of second language acquisition have mostly focused on investigating the negative effects of cross-linguistic influence, i.e. how the influence of the native language can lead to learning challenges and how to overcome them. By contrast, considerably less attention has been paid to the potentially beneficial role of cross-linguistic influence (Bardovi-Harlig & Sprouse, 2017; McManus, 2022; Ringbom, 2007). In this chapter, I explore whether cross-linguistic syntactic similarity facilitates adults' initial L2 syntax acquisition. Previous research has examined the facilitative role of L1 influence in learning L1-L2 similar structures or morphosyntactic features. Most studies have looked at how these structures or features are processed or used at different stages of L2 acquisition, with learners' proficiency ranging from low to advanced (e.g. low*,* Bardovi-Harlig, 1997; C. Ellis et al., 2013; Tokowicz & MacWhinney, 2005 low and/or medium, Chang & Zheng, 2015; Izquierdo & Collins, 2008*;* medium and/or advanced*,* Díaz et al., 2016; Foucart & Frenck-Mestre, 2011; Zawiszewski et al., 2011). To my knowledge, only Tolentino and Tokowicz (2014) have investigated whether cross-linguistic syntactic similarity facilitates L2 syntax learning by complete beginner adult learners. In this chapter, I present two experiments that further explore this under-researched area, focusing on whether Spanish natives with no knowledge of Galician show a learning advantage for a Spanish-Galician similar structure over a dissimilar one. As mentioned in Chapter 1, various approaches to L2 acquisition, including the MOGUL framework, implicitly or explicitly assume that L1 syntax is present at the initial state of L2 acquisition and, whenever possible, will be used to parse L2 input. If this is not appropriate, the linguistic system will need to be adapted, simply for the purpose of processing the L2. In this light, I hypothesized that Spanish natives would evidence a learning advantage for the Spanish-Galician similar structure, which could be processed using an L1 structure, over the Spanish-Galician dissimilar structure, which would have to be learnt from input. This hypothesis was met, although not conclusively, in Experiment 1. After some changes to the experimental design, it was fully met in Experiment 2.

Chapter 2 is organized as follows. In Section 1.1, I review some of the most relevant studies on the facilitative role of cross-linguistic syntactic similarity at the early stages of L2 syntax learning. After that, in Section 1.2, I describe the structures that constitute the object of study of this chapter. Then, in Section 1.3, I detail my hypotheses regarding how L1 syntax could affect the initial acquisition of L1-L2 similar and dissimilar structures, based on MOGUL and other approaches to second language acquisition. In Section 2, I describe and discuss Experiment 1 and in Section 3, I describe and discuss Experiment 2. Finally, in Section 4 I present some concluding remarks.

1.1. The facilitative role of L1 syntax at the early stages of L2 syntax learning

Several theories and models of second language acquisition assume, as advanced, that the L1 is the starting point for L2 acquisition and that this will facilitate the acquisition of cross-linguistically similar syntactic aspects, as opposed to different ones. Evidence in favour of this claim comes from a number of studies conducted over the last decades examining how adults at early stages of L2 acquisition process (morpho)syntactic structures or features with and without equivalents in the L1. Crucially, most studies tested learners who, while being at the early stages of L2 acquisition, were not complete beginners. In what follows, I review some of these investigations.

The first studies date back to the 1990s. Vainikka and Young-Scholten (1994) found that Turkish (n = 11) and Korean (n = 6) learners of L2 German having received "very little or no formal instruction" in the L2 (p.272) acquired a syntactic property shared in the L1 and the L2 (head-final VP) more easily than a property exclusive to the L2 (verb raising to main-clause second position). Learners conducted a series of tasks aiming to elicit sentences with diverse verb placements (e.g. tell the story in a comic strip, describe a set of pictures, describe a task that the experimenter performed or mentioned). Overall, all learners produced over 60% of verb-final sentences in obligatory contexts, which, according to Vainikka and Young-Scholten, indicated that they had acquired the head-final VP property of German. By contrast, 53% of learners showed no evidence of having acquired verb raising, i.e. they did not produce over 60% of sentences with a raised verb in main clauses. This suggests that the existence of a head-final VP property in Turkish and Korean facilitated its acquisition and use in the L2, but the fact that learners' L1 did not allow verb raising made its acquisition more challenging.

A few years after Vainikka and Young-Scholten's (1994) study, Bardovi-Harlig (1997) showed that L2 learners with different L1s used a grammatical form more appropriately when it was present in their L1 than when it was absent. The author conducted a 10-month-long study with native speakers of Spanish (n = 3), Japanese (n = 6), Korean (n = 2) and Arabic (n = 6), all enrolled in a beginning course in *English as a Second Language.* The study examined the timing of emergence of the present perfect and the effect of instruction in its use by L2 English learners. Spanish has a tense-aspect form that closely corresponds to the present perfect, but Japanese, Korean and Arabic have no grammatical equivalent. To assess learning of the present perfect, Bardovi-Harlig analysed learners' written texts throughout the course. Overall, there were 502 occurrences of the present perfect. Spanish natives produced the most instances (44.02%, 221/502), followed by the Japanese (37.25%, 187/502), the Arabic (13.75%, 69/502) and the Korean (4.98%, 25/502). In addition, the Spanish used the present perfect appropriately the most (87.78%, 194/221). The Japanese, the Arabic and the Korean used it as required on 74.87% (140/187), 76.81% (59/63) and 76% (19/25) of the occasions, respectively. These results suggest that the Spanish acquired the present perfect best, which could be attributed to the existence of a present perfect equivalent in their L1.

In the 21st century, the influence of L1-L2 similarities and differences during the early stages of L2 syntax acquisition continued to be a topic of interest. In line with Bardovi-Harlig's (1997) results, Izquierdo and Collins (2008) found that two French verb forms used to mark the perfective and the imperfective aspects —the *passé composé* and the *imparfait* —were more successfully acquired and used when these existed in learners' L1 than when they did not. Participants were a group of Mexican Spanish natives (n = 17) and a group of English natives (n = 15) with low-intermediate proficiency in L2 French. While Spanish has two verb forms equivalent to the *passé composé* and the *imparfait* to express the perfective/imperfective distinction (e.g. *yo sembré – j'ai planté* [perfective], *yo sembraba – je plantais* [imperfective]), both translate into the simple past in English (*I planted*). Participants conducted a cloze passage test with four verb types that can be used in perfective and imperfective contexts, namely stative (e.g. *être*, "to be"), activity (*marcher,* "to walk"), accomplishment (e.g. *nager,* "to swim") and achievement verbs (e.g. *tomber,* "to fall"). While stative verbs are prototypically used in imperfective contexts, the rest of verbs are most often used in perfective contexts. Overall, the Anglophones and the Hispanophones used the perfective comparably well, but the Hispanophones used the imperfective significantly more appropriately than the Anglophones. In addition, the Anglophones made more mistakes than the Hispanophones when verbs typically used in perfective contexts appeared in imperfective contexts. The authors argued that the existence of two verb forms to express the perfective/imperfective distinction in Spanish facilitated the acquisition of the equivalent verb forms in French.

L2 beginning learners have also been found to use a morphosyntactic feature more appropriately when it is present in the L1 than when it is absent, even when the L1 feature is not congruent with the L2 feature. For instance, C. Ellis et al. (2013) investigated how L1 Afrikaans (n = 23), English (n = 9) and Italian (n = 6) speakers having taken a beginning course in *German as a Second Language* acquired gender agreement between the determiner and the adjective. Italian and German both exhibit gender agreement between these elements, but English and Afrikaans do not. Importantly, while Italian has masculine and feminine gender, German has masculine, feminine and neuter gender. Thus, there is not a one-to-one correspondence between masculine/feminine determiners and adjectives in the two languages. To test gender agreement, participants conducted a sentence-completion task. First, they saw a sentence in German containing a noun with a masculine, feminine or neuter gender (e.g. *Das Mädchen ist schön* ('the-NEUTER girl-NEUTER is beautiful') p.21). This was followed by another sentence containing the same noun modified by a determiner and an adjective in Italian, Afrikaans or English between parenthesis (e.g. *Ich liebe (a beautiful) __________ Mädchen* ('I love (a beautiful) ________ girl.') p.21). Participants had to provide the correct translation of the determiner and the adjective into German. Overall, the Italian natives performed the task more accurately than the other two L1 groups. The fact that the Italian outperformed the Afrikaans and the English indicated that the first were more successful acquiring the German grammatical gender system. This was attributed to Italian and German sharing an abstract gender feature not present in English and in Afrikaans.

Chang and Zheng (2015) provided additional evidence supporting the facilitative role of L1-L2 similarities in L2 syntax learning. Their focus was on English and Chinese causatives, which can be expressed using syntactic, morphological and lexical constructions. Syntactic causative constructions are formed similarly in the two languages (they are periphrastic constructions including *make*-type verbs). Morphological and lexical causatives are formed differently across languages, with the exception of morphological structures with affixed causative verbs. Forty-five English natives enrolled in a beginning (n = 20) or intermediate (n = 25) Chinese course translated 12 sentences from their L1 to the L2. These included (i) two periphrastic constructions formed similarly in English and Chinese, (ii) two affixed causative constructions, also similar in English and Chinese, and (iii) eight morphological and lexical constructions expressed differently in the two languages. The two groups of learners translated English periphrastic causatives using the corresponding Chinese periphrastic structure, suggesting that the fact that this construction was similar in the two languages facilitated its learning. Beginning learners did not use Chinese morphological and lexical constructions. Instead, they overgeneralized the Chinese periphrastic constructions and used them to translate all types of causatives, even when this was not possible in Chinese. Intermediate learners used the cross-linguistically dissimilar morphological and lexical constructions, suggesting that these developed at later stages of L2 acquisition.

So far, I have discussed production studies, in which L2 syntax acquisition was assessed by looking at learners' spoken or written speech. Yet, some comprehension studies have also examined the positive influence of L1 syntax at early stages of L2 development. I will briefly mention a couple of them: Tokowicz and MacWhinney (2005) and Tolentino and Tokowicz (2014). Tokowicz and MacWhinney tested 34 English university students in a beginning semester in L2 Spanish on their sensitivity to three features: (i) progressive tense marking, formed similarly in English and Spanish, (ii) determiner-noun number agreement, formed differently in the two languages and (iii) determiner-noun gender agreement, unique to Spanish. Sensitivity was assessed in a Grammaticality Judgement Task (GJT) while Event-Related Potentials (ERP) were recorded. The ERP analysis indicated sensitivity to violations of the similar and the unique features, suggesting that these features had been learnt, but no sensitivity to violations of the dissimilar feature. Learners' ability to detect violations of the similar feature was attributed to the existence of an equivalent L1 feature. By contrast, their difficulty detecting violations of the dissimilar feature was ascribed to a competing L1 feature being active in learners' minds and hindering acquisition.

Following Tokowicz and MacWhinney (2005), Tolentino and Tokowicz (2014) looked at beginning L2 learners' sensitivity to three morphosyntactic features: one similar in the L1 and the L2, one dissimilar in the two languages and one unique to the L2. Tolentino and Tokowicz's participants were complete beginners and, thus, learnt the L2 features in the laboratory. Specifically, 39 English natives with no knowledge of Swedish were taught: (i) number agreement between demonstrative determiners and nouns (similar in the L1 and the L2), (ii) definiteness marking in noun phrases (dissimilar in the L1 and the L2) and (iii) gender agreement between the definite article and the adjective (unique to the L2). Participants conducted a longitudinal study that comprised four testing sessions[1]. In the first session, they were taught the Swedish words used in the experiment (articles, nouns, verbs, adjectives and adverbs). Next, they were exposed to pairs of grammatical sentences exemplifying the target morphosyntactic features and were encouraged to look for grammatical patterns in the sentences. In the second session, participants were trained on the vocabulary again and they were tested on their knowledge by means of a translation test (from L2 to L1). Then, they were

[1] Participants were pseudo-randomly assigned to one of three instruction groups: a Control group, a Salience group and a Rule and Salience group. Each group learnt the morphosyntactic features in slightly different ways. Since the relevance of this study for my dissertation resides in the investigation of initial L2 morphosyntax learning, I do not discuss the effectiveness of the three instruction methods.

exposed to the three morphosyntactic features one more time. Grammar learning was tested in a GJT with feedback conducted in sessions 2, 3 and 4. In the fourth and last session, participants were also trained on L2 morphosyntax prior to conducting the GJT. Tolentino and Tokowicz calculated *d'* scores for each morphosyntactic feature in each GJT[2]. Overall, *d'* scores for the similar and the unique features were significantly higher than for the dissimilar feature. This suggests that learning was significantly greater for the first two features than for the last one. The authors attributed the learning advantage of the similar feature over the dissimilar one to the influence of the L1, which facilitated acquisition of the former and hampered acquisition of the latter.

In sum, there is evidence that, at the early stages of L2 acquisition, the influence of the native language may facilitate the acquisition of cross-linguistically similar L2 structures and morphosyntactic features compared to cross-linguistically dissimilar ones. Most research has tested learners who had been exposed to the L2 prior to the study. For instance, participants in all but two of the studies reviewed (Tolentino & Tokowicz, 2014 and Vainikka & Young-Scholten, 1994) had completed a course in the L2 or were currently enrolled in one[3]. In addition, the amount of exposure leading to L2 acquisition and/or the conditions under which this exposure takes place are often not controlled for. Within the review presented, the two or more groups of participants in a study sometimes completed the L2 course at different universities and, hence, under different circumstances. For example, C. Ellis et al.'s (2013) Afrikaans and English natives had taken a German course in the same South-African university, but Italian natives had taken a German course in an Italian university. Likewise, Izquierdo and Collins' (2008) English natives were taking French classes in Canada, but Spanish natives were taking French classes in Mexico. On some occasions, participants were also exposed to the L2 outside of the classroom. Vainikka and Young-Scholten (1994) tested participants who had been living in Germany for a number of years (range 1.5-24), a period during which they had naturally been exposed to German. Bardovi-Harlig's (1997) participants lived in an English-speaking country, so they had been exposed to English prior to entering the L2 course and were exposed to English outside of the class throughout the longitudinal study. Chang and Zheng's (2015) participants were studying a Chinese major in Shanghai, so they lived in a Chinese-dominant environment. Finally, Izquierdo and Collins' (2008) Anglophone participants studied in Montreal, so they probably had some contact with French in their daily lives. All this makes it hard to know if the exposure leading to acquisition was comparable across learners and syntactic phenomena and should be taken into account when interpreting the results of the aforementioned studies. In addition, some of the studies, notably Vainikka and Young-Scholten (1994), Bardovi-Harlig (1997), Izquierdo and Collins (2008) and C. Ellis et al. (2013) were conducted with a small number of participants, and the groups of learners

---

[2] The *d'* score reflects sensitivity to the difference between grammatical and ungrammatical sentences unaffected by response bias (Stanislaw & Todorov, 1999).

[3] Vainikka and Young-Scholten reported that their participants had received "very little or no formal instruction in German" prior to testing, but they did not specify what "very little" means.

compared had unequal sample sizes. The results of these studies could usefully be replicated with a sample size that is larger and comparable across groups, which would be more representative of the general population.

Due to learners' previous knowledge of the second language, studies examining cross-linguistic influence at the early stages of L2 syntax learning often do not involve in-lab teaching. What is tested is, essentially, low proficient L2 learners' knowledge of syntactic structures or features acquired in a context not controlled by the researcher. To my knowledge, Tolentino and Tokowicz's (2014) study is the only one that has explored how adult second language learners without any previous knowledge of the L2 learn syntactic features differing in similarity with the L1, with learning happening in the lab and controlling for the amount and the type of exposure that participants receive. Chapter 2 aims to shed some more light on this under-researched topic. In what follows, I introduce the cross-linguistically similar and dissimilar structures studied in this chapter.

1.2. Cross-linguistically similar and dissimilar structures studied in Chapter 2

The object of study of Chapter 2 are two Galician structures differing in similarity with Spanish. These are an embedded clause with a subjunctive verb (cross-linguistically similar structure, 1a) and an embedded clause with an inflected infinitive verb, i.e. a verb that agrees in person and number with its subject (cross-linguistically dissimilar structure, 1b).

(1)	a.	É	importante	[que	Pedro	repare	a	radio].
		it.is	important	COMP	Pedro	fix $_{PRS.SBJV.3SG}$	the	radio
		"It is important that Pedro fixes the radio."						
	b.	É	importante	[Pedro	reparar	a	radio].	
		it.is	important	Pedro	fix $_{INF[AGR.3SG]}$	the	radio	
		"It is important that Pedro fixes the radio."						

(Example sentences used in Experiment 1)

The sentences in (1) express the same meaning. They both start with an impersonal expression conveying opinion (*É importante,* "It is important") followed by an embedded clause (between brackets). In (1a), this clause is introduced by a complementizer (*que,* "that") and contains a verb in the present subjunctive. In (1b), the embedded clause is not introduced by a complementizer and contains an inflected infinitive verb. The subject of the embedded clause is a third person singular noun (*Pedro*); in this case, the inflected infinitive does not overtly mark subject-verb agreement[4]. The alternation in (1) is productive in Galician, as shown in (2), where the elided subject of the embedded clause is a third person plural:

[4] See the grammar from the Real Academia Galega and Instituto da Lingua Galega (2012) for the full paradigm of an inflected infinitive verb.

(2) a. É doado [que supoñan as cousas].
it.is easy COMP suppose$_{PRS.SBJV.3PL}$ the things
"It is easy that they suppose the things."

b. É doado [supoñeren as cousas].
it.is easy suppose$_{INF-AGR.3PL}$ the things
"It is easy that they suppose the things."

(Longa, 1994, p. 25)

As shown in (1), the embedded structures studied in this chapter had a pre-verbal subject. In Galician embedded clauses with inflected infinitives, the subject, if overtly realized, must follow the infinitive, as indicated in (3):

(3) a. Será difícil [aprobaren eles a proposta].
it.will.be difficult approve$_{INF-AGR.3PL}$ they the proposal
"It will be difficult that they approve the proposal."

b. *Será difícil [eles aprobaren a proposta].
it.will.be difficult they approve$_{INF-AGR.3PL}$ the proposal
"It will be difficult that they approve the proposal."

(Sheehan & Parafita Couto, 2011, p. 2)

Nevertheless, as exemplified in (4), pre-verbal subjects in this type of clauses exist in European Portuguese, a language closely related to Galician:

(4) a. Será difícil [eles aprovarem a proposta].
it.will.be difficult they approve$_{INF-AGR.3PL}$ the proposal
"It will be difficult for them to accept the proposal."

b. *Será difícil [aprovarem eles a proposta].
it.will.be difficult approve$_{INF-AGR.3PL}$ they the proposal
"It will be difficult for them to accept the proposal."

(Raposo, 1987, p. 86)

Thus, the inflected infinitive clause in (1b) is attested in a natural language, i.e. it is not a structure from an impossible human language (cf. Kallini et al., 2024; Mitchell & Bowers, 2020). The modification of the Galician post-verbal subject to a pre-verbal subject aimed to make the infinitive clause more comprehensible for Spanish natives, given that Spanish is an SVO language (López, 1997).

Like Galician, Spanish admits embedded clauses introduced by *que* and followed by a verb in the present subjunctive, as exemplified in (5a), the Spanish equivalent of (1a). It also admits embedded clauses without the complementizer and introduced by an infinitive verb (5b). Importantly, in (5b) the verb is a non-inflected infinitive, for Spanish does not have inflected infinitives, and the non-finite embedded clause does not admit an overt subject (Bosque &

Gutiérrez-Rexach, 2009). Hence, the inflected infinitive construction in (1b) would be ungrammatical in Spanish.

(5) a. Es importante [que Pedro repare la radio].
it.is important COMP Pedro fix$_{\text{PRS.SBJV.3SG}}$ the radio
"It is important that Pedro fixes the radio."

b. Es importante [(*Pedro) reparar la radio].
it.is important Pedro fix$_{\text{INF}}$ the radio
"It is important that Pedro fixes the radio."

1.3. The facilitative role of L1 syntax in initial L2 syntax learning

In Chapter 1, I briefly mentioned how L1 syntax could facilitate the acquisition of cross-linguistically similar L2 structures, as opposed to cross-linguistically dissimilar L2 structures, within several models and theories of L2 acquisition (Section 2). I also proposed how this could take place within the MOGUL framework (Section 3.5). In what follows, I review and expand these claims. I discuss the acquisition of cross-linguistically similar structures first (Section 1.3.1) and, then, I turn to the acquisition of cross-linguistically dissimilar ones (Section 1.3.2).

1.3.1. The acquisition of cross-linguistically similar L2 structures

As discussed in Chapter 1, MOGUL describes language in the mind in terms of a *core language system* (consisting of a phonological and a syntactic module) and *language broadly defined* (including auditory, visual, articulatory and conceptual modules, which interact with the core language system during language processing, see Figure 1.1). Each module has an information store that contains (acoustic, orthographic phonological, syntactic or conceptual) representations of linguistic knowledge and a processor that selects and combines these representations during input processing. Interfaces activate and coindex representations in adjacent modules to form chains of representations (see Figure 1.2). A syntactic structure is a syntactic representation formed by combining smaller syntactic representations and is coindexed with the phonological and conceptual representations of particular lexical items as encountered in the input (Sharwood Smith & Truscott, 2014, sec. 2.3). Regarding L2 acquisition, MOGUL assumes that this starts with all the representations and chains of representations established during L1 acquisition in the information stores. Since these representations have been extensively processed as part of the L1, they have high resting activation levels (Sharwood Smith & Truscott, 2014, sec. 10.3). From this assumption, I infer that the initial state of L2 acquisition includes all the syntactic structures of the L1 stored with a high resting activation level. In addition, MOGUL understands language acquisition as the result of processing, and processing is considered essentially cross-linguistic. That is, Sharwood Smith and Truscott assume that the processing system is shared for the L1 and the L2 and that items from all languages in the mind activate during acquisition and processing (Sharwood Smith & Truscott, 2014, sec. 6.5.2). For instance, the authors mention that processing an L2 word for the first time requires associating its phonological form with a syntactic (and a conceptual) representation and that, upon receiving input from the phonological module, compatible L1 syntactic representations activate in the syntactic store. One of these representations is selected by the syntactic processor and associated with the L2 word. If the representation is not problematic, a novel representation will not be created (Sharwood Smith & Truscott, 2014, sec. 4.2, 7.4.2). Considering all this, I infer that, during L2 processing, L1 syntactic structures will activate and that, when exposed to a novel L2 structure, the syntax processor will try to process it using a compatible L1 representation.

Taking all this into account, I propose that the first time that an L2 learner is aurally or visually exposed to a sentence formed by a structure shared in the L1 and the L2 (for example, the cross-linguistically similar structure in 1a, Section 1.2), an auditory or visual representation of it will incrementally activate in the auditory or visual stores. As input is received, a suitable phonological representation will activate through the interface between the (visual-) auditory and the phonological modules and the two representations will be coindexed. At the same time, the interface between the phonological and the syntactic modules will incrementally activate a syntactic representation compatible with it. This will be the representation of the cross-linguistically similar structure acquired as part of the L1. This syntactic representation, in turn, will incrementally activate conceptual representations coindexed with it. The index of the phonological representation of the L2 structure will be assigned to the syntactic and conceptual representations active. Simply put, the L2 structure will be processed as if it was equivalent to its L1 counterpart, using a syntactic representation already available in the learner's linguistic system. Importantly, the structure shared in the L1 and the L2 will have a high resting activation level. This means that, at the initial stage of L2 acquisition, it will already be firmly established in the linguistic system. Each time that the structure is processed, its resting activation level will increase.

The processing of a cross-linguistically similar L2 structure within MOGUL that I have proposed is broadly compatible with theories and models of L2 acquisition such as B.D. Schwartz and Sprouse's Full Transfer/Full Access model (FT/FA model, 1994, 1996), Carroll's Autonomous Induction Theory (AIT, 1999, 2001), MacWhinney's Unified Competition Model (UCM, 2005) and Westergaard's Micro-cue model of L2 acquisition (2021). The **FT/FA model** assumes that "the initial state of L2 acquisition is the final state of L1 acquisition" and that "the entirety of the L1 grammar […] immediately carries over as the initial state of a new grammatical system" (B. D. Schwartz & Sprouse, 1996, pp. 40–41). When the learner parses an L2 structure for the first time, it does so by recurring to this new grammatical system, which is no different from L1 grammar. Thus, a structure that is similar in the L1 and the L2 will be processed without difficulty using a compatible L1 structure. Carroll's **AIT** also assumes that the whole of L1 grammar is present at the initial state of L2 acquisition and that during initial L2 syntax acquisition, the parsing system will attempt to process L2 input using the existing L1 grammar. When the construction to be processed is similar in the L1 and the L2, processing will proceed smoothly. The **UCM** is an extension of the Competition Model, first postulated as a functionalist approach to first language acquisition and processing (E. Bates & MacWhinney, 1981, 1982; MacWhinney, 1987a) and later applied to second language learning (MacWhinney, 1987b, 1992, 1997). The UCM conceives language acquisition and processing as the process of mapping forms and functions. To interpret the meaning of a sentence, individuals pay attention to linguistic forms, which contain phonological and syntactic information seen as *cues* activating different functions or meanings. L2 syntax learning is seen as a cue acquisition process, which initially relies on cues from the L1. When the linguistic cues are similar in the L1 and the L2, applying L1 cues to the L2 will be positive. Finally, the **Micro-cue model of L2 acquisition** assumes that the L1 is part of the initial state of the L2, that language learning results from processing (and, hence, involves no learning-specific mechanisms) and that all languages a speaker knows are active and can be accessed during processing. Grammar is conceived as a set of micro-cues or micro-structures that learners extract from the input[5]. When learners are exposed to the L2 for the first time, they try to parse it according to the L1's micro-cues and, if the structure (micro-cue) to be processed has an identical L1 counterpart, processing is facilitated.

1.3.2. The acquisition of cross-linguistically dissimilar L2 structures

In Chapter 1, I mentioned how, within MOGUL's *Acquisition by Processing Theory,* if none of the (phonological, syntactic or conceptual) representations in the linguistic system is appropriate to process the input, a new representation will be constructed. This newly created representation will have a low resting activation level. If it is never used in processing again, this activation will disappear and the representation will not be stored or acquired. Conversely, if the representation is used to process subsequent input, its resting activation level will increase and it will gradually become part of the linguistic system (Sharwood Smith & Truscott, 2014, sec. 4.2). From this information, I infer that acquiring a cross-linguistically dissimilar L2 structure will require creating and storing a novel syntactic representation in the syntactic module.

I propose that when an L2 learner is first exposed to a sentence formed by a structure that is different from its L1 equivalent (for example, the cross-linguistically dissimilar structure in 1b, Section 1.2) processing will initially be like that described for the cross-linguistically similar structure. That is, when the learner hears or reads the sentence, an auditory/visual and phonological representation of that sentence will incrementally activate in the auditory/visual and phonological stores. As this occurs, the interface between the phonological and the syntactic modules will try to coindex these representations with a compatible syntactic representation in the syntactic store. Crucially, no available (L1) representation will be adequate. In this case, the syntactic processor will have to select active items in the syntactic store and combine them to construct an appropriate representation of the input. At the same time, the conceptual processor will have to select a likely conceptual representation, which may or may not be correct. The newly created syntactic representation will have a very low

[5] For example, in Norwegian, micro-cues capturing in which contexts questions require a V2 word order, a non-V2 word order, or allow both word orders (Westergaard, 2021).

resting activation level. Each time that the structure is processed as part of the L2, its resting activation level will increase and it will gradually become a stable item in the linguistic system. If processing opportunities are not enough, the resting activation level of this structure might not be sufficiently high for it to become part of the grammar. This contrasts with the situation discussed for the cross-linguistically similar structure, which was assumed to be a very stable item already at the initial stage of L2 acquisition.

Overall, the acquisition of a cross-linguistically dissimilar L2 structure within MOGUL, as I have proposed it, is compatible with the approaches to L2 acquisition mentioned in the previous section. In the **FT/FA model,** when the structure to be processed is different in the L1 and the L2, the L2 grammar needs to be restructured. This is argued to occur by accessing UG. The **AIT** suggests that if the construction to be processed is different in the L1 and the L2, the parsing mechanisms will not be able to process it according to the information in the linguistic system and the learning mechanisms will trigger the restructuring of this information to account for the novel input. In terms of the **UCM,** when exposed to a cross-linguistically dissimilar L2 structure, learners will not be able to process the input using L1 cues and will have to acquire the appropriate L2 cue. Finally, in the **Micro-cue model of L2 acquisition,** if an L2 structure cannot be successfully parsed using the L1's micro-cues, as it would occur for L1-L2 dissimilar structures, learners must postulate new micro-cues that are consistent with the target language.

2. Experiment 1

2.1. Overview

The research question of Experiment 1 was "Does cross-linguistic syntactic similarity facilitate initial L2 syntax acquisition?" To address this question, I exposed Spanish natives with no knowledge of Galician to the cross-linguistically similar and dissimilar L2 structures presented in Section 1.2. Then, I examined and compared the establishment of the two constructions in learners' linguistic system (in familiar thinking terms, I examined and compared learning of the two structures). I created a mini-language based on Galician that contained the two target structures and where all the vocabulary was cognate with Spanish. Cognates were used so that learners could process the structures without going through a vocabulary-learning phase. I hypothesized that, despite not knowing any Galician, learners would be able to process and comprehend all cognate words because these would simply be processed as if they were equivalent to their L1 counterparts. Experiment 1 used an implicit learning paradigm. Learners were exposed to the structures by means of a Plausibility Judgement Task (PJT). Then, learning of these structures was tested in a Grammaticality Judgement Task (GJT). The experiment ended with an awareness measure in the form of a verbal report. My hypotheses regarding the acquisition of the cross-linguistically similar and dissimilar L2 structures were already detailed in Section 1.3 and are summarised here. In short, based on the MOGUL framework and other approaches to L2 acquisition, I assumed that the cross-linguistically similar structure would be processed using an L1 structure present in learners' linguistic system from the beginning of L2 acquisition, but that the cross-linguistically dissimilar structure would need to be learnt from input. Considering this, I hypothesized that following exposure to the two structures, two scenarios could be in place. If the input was not enough to trigger learning of the cross-linguistically dissimilar structure, then **Hypothesis 1 (H1)** claimed that the cross-linguistically similar structure would be established in learners' linguistic system, but the cross-linguistically dissimilar structure would be not. Alternatively, if the input triggered learning of the cross-linguistically dissimilar structure, then **Hypothesis 2 (H2)** claimed that the cross-linguistically similar and dissimilar structures would both be established in learners' linguistic system, but that the former would be more firmly established than the latter.

2.2. Participants

Twenty-four Spanish natives (21 female) took part in the experiment. Their ages ranged from 18 to 40 (M = 23, SD = 5.18) and they were all students at the University of the Basque Country (UPV/EHU). All of them reported having no previous knowledge of Galician or other languages with inflected infinitives in a linguistic background questionnaire filled out prior to the experiment (adapted from Weber-Fox and Neville, 1996; see Appendix A-1). Spanish natives living in the Basque Country may be bilingual with Basque. Hence, the questionnaire asked participants about their language history, proficiency and use in both languages[6]. The majority of participants (79%, 19/24) knew Basque, but they all reported feeling more comfortable using Spanish. In addition, the 24 participants reported using only Spanish with their parents when they were little (0-3 years).

The linguistic background questionnaire asked about the frequency of use of Spanish and Basque in three periods of life: childhood (3-12 years), puberty (12-18 years) and adulthood (after 18 years) in three environments: at school/university/work, at home and at other places. Responses were scored on a 7-point scale in which 1 stood for *Spanish only* and 7 stood for *Basque only.* The mean scores for each environment in each life period can be found in Appendix A-2. In sum, the language participants currently used and had used the most in their lives was Spanish (mean language use during childhood, 1.46 (SD = 0.96); puberty, 1.44 (SD = 0.90); adulthood, 1.29 (SD = 0.62)). The questionnaire also asked about self-rated proficiency in Spanish in four skills: speaking, listening, reading and writing. Responses were scored on a 7-point scale in which 1 represented *very poor proficiency* and 7 represented *perfect proficiency.* The mean scores for each skill are reported in Appendix A-2. Overall, participants considered they had a nearly perfect mastery of Spanish (mean proficiency level collapsing all skills, 6.88 (SD = 0.39)).

---

[6] In Basque, the meaning expressed by an embedded clause with a subjunctive verb may also be conveyed by an embedded clause with a nominalized infinitive (Hualde & Ortiz de Urbina, 2003). However, these are not the same as Galician inflected infinitives. In this light, I considered that the fact that participants had knowledge of Basque was not an impediment for them to participate in the study. Learners' proficiency level in Basque was at B2 or under.

All participants reported having normal or corrected to normal vision and hearing. Before the experiment began, they read and signed an informed consent (Appendix A-3). This experiment was part of the project "Cross-linguistic activation effects in bilingual language processing and learning" (PGC2018-097970-B-I00), funded by the Spanish Ministry of Science, Innovation and Universities and approved by the Committee of Ethics for research involving human beings of the University of the Basque Country (*Comité de Ética para las Investigaciones con Seres Humanos, CEISH,* Ref. M10_2019_167). Participants were paid 7€ for their participation.

2.3. Materials

As mentioned in Section 2.1, a mini-language consisting of Spanish-Galician cognates and Galician-based syntax was used to generate the stimuli for this experiment. The object of study were the two structures differing in cross-linguistic similarity between Spanish and Galician described and exemplified in Section 1.2. Here, I review and further detail their characteristics. The two structures consisted of an embedded clause attached to an impersonal expression conveying opinion (e.g. *É importante,* "it is important") and expressed equivalent meanings. In both cases, the embedded clause acted as the subject of the structure. One of the structures exists in Spanish and Galician and will henceforth be referred to as *similar structure.* The other structure only exists in Galician (even if the structure in this experiment was slightly adapted from Galician in that the typically post-verbal subject was made pre-verbal). Since the meaning expressed by this structure would be expressed with a different structure in Spanish (participants' L1), it will be referred to as *dissimilar structure* (Table 2.1).

Structure	**Example**
Similar structure	É importante [que Pedro repare a radio] it.is important COMP Pedro fix $_{\text{PRS.SBJV.3SG}}$ the radio "It is important that Pedro fixes the radio."
Dissimilar structure	É importante [Pedro reparar a radio] it.is important Pedro fix $_{\text{INF[AGR.3SG]}}$ the radio "It is important that Pedro fixes the radio."

TABLE 2.1. Examples of the similar and the dissimilar structure studied in Experiment 1. Embedded clauses are between brackets.

The difference between the similar and the dissimilar structure resides in their embedded clauses. In the similar structure, the embedded clause is introduced by the complementizer *que* ("that") and is followed by a subject in the form of a proper noun, a transitive verb in the present subjunctive and an article plus a singular inanimate noun acting as direct object. The dissimilar structure differs from the similar structure in that its embedded clause is not introduced by *que* and its embedded transitive verb is not conjugated in the present subjunctive, but it is an inflected infinitive. In both structures, the embedded verb agrees with a third person singular subject. In this case, the Galician inflected infinitive does not bear an overt person and number mark and is formally equivalent to the non-inflected infinitive. These two structures were used to generate a set of sentences for the exposure phase (henceforth, the *exposure set*) and the testing phase (henceforth, the *testing set,* additionally including ungrammatical sentences violating these structures). Sentences were created using the online automatic translator *Gaio*, a tool made available by the General Secretary for language policy of the Galician government (*Secretaría xeral de política lingüística*, *Xunta de Galicia*). Then, a native speaker of Galician checked that the vocabulary, verb conjugation and use of articles in the sentences was correct. Because sentences would be aurally presented in the exposure and testing phases, they were recorded by a female native speaker of Galician. Recordings took place in a soundproof booth using an Olympus voice recorder (Linear PCM Recorder LS-5 model, frequency sampling of 96kHz). The speaker read the sentences at a normal pace and with natural intonation. I cut the initial and final silences in each recording using *Praat* (Boersma & Weenink, 2018, version 6.0.37). The exposure set and the testing set are described in the next sections.

2.3.1. Exposure set

The exposure phase consisted of a plausibility judgement task. Hence, the exposure set contained plausible and implausible sentences. First, I created 100 plausible sentences formed by the similar structure. Next, I generated three other versions of each sentence manipulating cross-linguistic similarity (Similar structure vs. Dissimilar structure) and plausibility (Plausible construction vs. Implausible construction) (see Table 2.2 for examples of the four conditions). Plausibility was manipulated by varying the direct object that followed the transitive verb in each sentence. Thus, the last word of the sentence indicated the plausibility of the proposition expressed. This manipulation ensured that learners could only judge if the sentence was plausible or not after processing the whole construction. The four versions of each sentence were divided into four different lists. Each list contained 100 items (25 plausible sentences and 25 implausible sentences formed by the similar structure, 25 plausible sentences and 25 implausible sentences formed by the dissimilar structure).

Before deciding on the final set of sentences that would constitute the exposure set, I conducted a norming study to test that my plausibility judgements coincided with those of the target participants (Spanish natives with no knowledge of Galician). The norming study consisted of a plausibility judgement task and tested only the 100 plausible and implausible sentences formed by the similar structure. Sentences were presented in Spanish. To prevent participants from seeing both versions of the same sentence, these were divided into two lists of 100 items, each containing 50 plausible constructions and 50 implausible ones. Twenty Spanish natives (11 men, M_{age} = 24.65, *SD* = 4.13, age range = 19-33) recruited via social media participated in the study. Participants were considered native speakers of Spanish if they reported having been born and raised in Spain and speaking only Spanish at home. The plausibility judgement task was administered through the online platform Ibex Farm (Drummond, 2013). Participants were informed that they would see some sentences in Spanish and would have to judge their plausibility. They were told to judge a sentence as *Plausible* if they considered it was “valid, coherent and made sense” and as *Implausible* otherwise. The instructions were accompanied by two example sentences —one plausible and one implausible— not included in the experimental items. Sentences were presented one at a time and in a randomized order for each participant. Each sentence appeared in the middle of a white screen with the options *Plausible* and *Implausible* written below. The key "A" had to be pressed to select the option *Plausible* and the key "L" to select the option *Implausible.* There was no time limit to respond. Participants' mean accuracy percentage was 97.40% (*SD* = 15.92%, 95%CI = [96.41, 98.39]) for plausible sentences and 91.90% (*SD* = 27.30%, 95%CI = [90.21, 93.59]) for implausible ones. Three plausible constructions and 20 implausible ones were misjudged by two or more speakers. I changed those sentences and asked participants to re-evaluate the new 23 items. On this occasion, all participants agreed on the plausibility/implausibility of the stimuli.

Condition	**Example**
Plausible similar structure (Pl.SS)	É importante que Pedro repare a radio it.is important COMP Pedro fix $_{PRS.SBJV.3SG}$ the radio "It is important that Pedro fixes the radio."
Implausible similar structure (Impl.SS)	? É importante que Pedro repare o tornado it.is important COMP Pedro fix $_{PRS.SBJV.3SG}$ the tornado "It is important that Pedro fixes the tornado."
Plausible dissimilar structure (Pl.DS)	É importante Pedro reparar a radio it.is important Pedro fix $_{INF[AGR.3SG]}$ the radio "It is important that Pedro fixes the radio."
Implausible dissimilar structure (Impl.DS)	? É importante Pedro reparar o tornado it.is important Pedro fix $_{INF[AGR.3SG]}$ the tornado "It is important that Pedro fixes the tornado."

TABLE 2.2. Examples of the four experimental conditions of the exposure set in Experiment 1. All sentences are grammatically correct. Following the convention in linguistics, implausibility is indicated by a question mark.

The lexicon of the exposure set was made up of five impersonal expressions conveying opinion or emotion, 20 verbs and 60 nouns (see Appendix A-4 for the complete list of vocabulary). The five impersonal expressions were *É importante* ("it is important"), *É posible* ("it is possible"), *É probable* ("it is probable"), *É sorprendente* ("it is surprising") and *É necesario* ("it is necessary"). Twenty proper nouns (10 male, 10 female) existing in both Spanish and Galician acted as the subject of embedded clauses. Forty inanimate singular nouns (26 masculine, 14 feminine; $p > .05$), preceded by the article *a* ("the", feminine) or *o* ("the", masculine), acted as the direct object of the embedded verb. Twenty nouns were used in plausible sentences and the remaining 20, in implausible sentences. Nouns changed from plausible to implausible sentences because not all nouns in plausible conditions could be combined with the verbs of the exposure set to express implausible propositions. Each impersonal expression occurred five times in each of the two plausible conditions and five times in each of the two implausible conditions across lists. Proper nouns, verbs and inanimate nouns appeared once or twice in each of the four conditions across lists. These items were comparable regarding length (number of letters), frequency per million of Spanish translations (SUBTLEX-ESP, Cuetos et al., 2011)[7] and level of phonological overlap with their Spanish counterparts (Levenshtein distance)[8] across conditions in all lists (all p > .05, see Appendix A-5). Each impersonal expression occurred with a particular proper noun, verb and inanimate noun just once in the exposure set. This can be consulted in Appendix A-6.

2.3.2. Testing set

The testing phase consisted of a grammaticality judgement task. The testing set included, thus, grammatical and ungrammatical sentences. I generated 80 grammatical sentences formed by the similar structure. Then, I created three additional versions of each sentence: a grammatical sentence formed by the dissimilar structure, an ungrammatical sentence violating the similar structure and an ungrammatical sentence violating the dissimilar structure. Grammaticality violations occurred in the embedded verb. In ungrammatical sentences violating the similar structure, the embedded verb was in the infinitive form. In ungrammatical sentences violating the dissimilar structure, the embedded verb was in the present subjunctive (see Table 2.3 for examples of the four conditions). I hypothesized that learners would process a sentence violating the similar or the dissimilar structure as if it was its grammatical counterpart up to the moment the verb was perceived, when the ungrammaticality would be detected. The four versions of each sentence were split into four different lists. Each list contained 80 items (20 grammatical sentences formed by the similar structure and 20 ungrammatical sentences violating it, 20 grammatical sentences formed by the dissimilar structure and 20 ungrammatical sentences violating it).

[7] Learners had never been exposed to Galician so, to them, Galician lexical items had no frequency of use. Yet, since I hypothesized that learners would process cognates as if they were equivalent to their Spanish counterparts, I considered it relevant that the Spanish translations of all cognates were matched in frequency.

[8] Levenshtein distance measures the number of insertions, deletions or substitutions needed to transform one word into another. For instance, the phonological Levenshtein distance between the English word *correct* (/kəˈrekt/) and its Spanish translation *correcto* (/koˈrekto/) is 2, since to transform the first word into the second it is necessary to substitute the first vowel of the word and to insert a vowel at the end. In Experiment 1, the Levenshtein distance between Galician and Spanish words was calculated with the assistance of a native speaker of Galician.

Condition	Example[9]
Grammatical similar structure (SS)	É importante que Antonio firme a carta it.is important COMP Antonio sign $_{\text{PRS.SBJV.3SG}}$ the letter "It is important that Antonio signs the letter."
Ungrammatical similar structure (*SS)	*É importante que Antonio firmar a carta it.is important COMP Antonio sign $_{\text{INF[AGR.3SG]}}$ the letter "It is important that Antonio signs the letter."
Grammatical dissimilar structure (DS)	É importante Antonio firmar a carta it.is important Antonio sign $_{\text{INF[AGR.3SG]}}$ the letter "It is important that Antonio signs the letter."
Ungrammatical dissimilar structure (*DS)	*É importante Antonio firme a carta it.is important Antonio sign $_{\text{PRS.SBJV.3SG}}$ the letter "It is important that Antonio signs the letter."

TABLE 2.3. Examples of the four experimental conditions of the testing set in Experiment 1. Following the convention in linguistics, ungrammaticality is indicated by an asterisk.

The lexicon of the testing set was made up of the same five impersonal expressions used in the exposure set, 20 novel verbs and 40 novel nouns (see Appendix A-4 for the complete list of vocabulary). I chose novel nouns and verbs for the testing set, instead of keeping those in the exposure set, to avoid that participants judged a sentence as grammatical simply because they recalled that that particular combination of lexical items had appeared during the exposure phase. Using a novel vocabulary in the testing set is a common practice in artificial and natural grammar learning experiments, which, as Experiment 1, assume that participants learn abstract structures, not specific exemplars of the language (see Reber, 1969; more recently Kim & Fenn, 2020; Rebuschat, 2009; Rogers, Révész, et al., 2015). As in the exposure set, 20 proper nouns (10 male, 10 female) shared in Spanish and Galician functioned as the subject of the embedded clause. Twenty inanimate singular nouns (9 masculine, 11 feminine; $p > .05$) preceded by an article acted as the direct object of embedded verbs. Each impersonal expression occurred four times per condition across lists. Proper nouns, verbs and inanimate nouns occurred once per condition across lists. The combination of a given impersonal expression, proper noun, verb and inanimate noun occurred just once in the testing set. This can be found in Appendix A-6.

2.4. Procedure

Participants were told that they would take part in a study on sentence comprehension in Galician. The experiment consisted of three phases: an exposure phase (Section 2.4.1), a testing phase (Section 2.4.2) and a debriefing phase (Section 2.4.3). It was run on the E-Prime 2.0 software (Schneider et al., 2002). Participants were tested individually in a soundproof booth. All sentences were played through headphones. The experiment was conducted in

[9] I provide the intended English translation for all sentences irrespective of their grammaticality.

Spanish and lasted for a maximum of 45 minutes. The instructions can be found in Appendix A-7.

2.4.1. Exposure phase

Participants were told that they would listen to some sentences in Galician. They were instructed to pay attention to the meaning of each sentence and to judge whether it was "Plausible (that is, logical, which makes sense)" or "Not Plausible (that is, illogical, which does not make sense)". They were not informed that half of the sentences were formed by the similar structure and, the other half, by the dissimilar structure, nor that they would be tested on their syntax knowledge afterwards. Sentences were played one at a time and were presented in a randomized order for each participant. While a sentence played, participants looked at a black cross at the centre of a white computer screen. Immediately after a sentence ended, the black cross was replaced by the options *Plausible* ("Plausible") and *No plausible* ("Not plausible") and participants had to make their plausibility judgement. The key "A" had to be pressed to indicate that the sentence was plausible and the key "L" to indicate that it was not plausible. The response keys were shown on the screen, below the options they represented. The words *Plausible* and *No plausible* remained on the screen for 5 seconds. If after this time no choice was made, a message indicating that no response was detected appeared and the next sentence was automatically played. If one of the two options was chosen, feedback was provided regarding the accuracy of the response, in the form of a green tick (right answer) or a red cross (wrong answer). These were shown for 700ms, after which a new sentence was played (Figure 2.1). Prior to the task, participants performed a practice session to familiarize themselves with how to make plausibility judgements. They listened to four sentences (exemplifying the plausible similar structure, the plausible dissimilar structure, the implausible similar structure and the implausible dissimilar structure) not included in the exposure set. The exposure phase lasted for around 10 minutes.

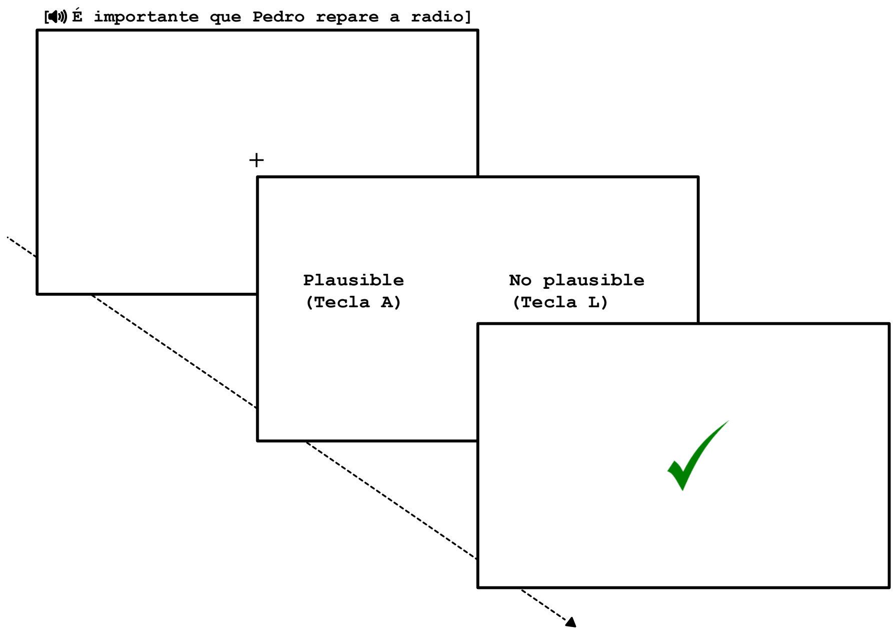


FIGURE 2.1. Example exposure trial in Experiment 1. First, participants listened to a sentence while looking at a cross in the middle of the screen. In the figure, this auditory sentence is presented between brackets and preceded by a speaker on top of the first screen. When the sentence ended, the options *Plausible* (“Plausible”) and *No Plausible* (“Not plausible”) appeared on the screen. Written below these options were the response keys, *Tecla A* (“Key A”) or *Tecla L* (“Key L”). After pressing one of the two, feedback was displayed for 700ms.

2.4.2. Testing phase

Once the exposure phase was completed, participants were informed that the sentences they had listened to “were formed according to two structures”. Then, they were told that they would listen to new sentences, all plausible. They were instructed to decide, as quickly as possible, whether the new sentences were formed by the same structures as the sentences in the previous task or not. To do this, they were told to use their intuition. Time pressure aimed to prevent participants from consciously reflecting about the form of the sentences, further encouraging them to use their intuition to perform the task. Learners were taught that those sentences that were formed according to the structures of the language were to be judged as “correct” and those that were not, as “incorrect”. Like in the exposure phase, participants listened to the sentences on an item-by-item basis while looking at a black cross located at the centre of a white computer screen. The presentation of the sentences was randomized for each participant. Once a sentence ended, the response options *Correcto* (“Correct”) and *Incorrecto* (“Incorrect”) appeared on the screen. The key “A” had to be pressed to indicate that the sentence was correct; the key “L” to indicate that it was incorrect. The response keys appeared below the options they represented. No feedback was provided regarding the accuracy of responses. The options *Correcto* and *Incorrecto* appeared on the screen for 5 seconds. If after this time participants had not chosen one of the two, a message saying that no response was detected appeared and a new sentence was played. The testing phase was preceded by a short practice session so that participants familiarized themselves with the new task. This included four sentences (exemplifying the grammatical similar structure, the ungrammatical similar structure, the grammatical dissimilar structure and the ungrammatical dissimilar structure) not part of the testing set.

2.4.3. Debriefing phase

The debriefing phase consisted of a verbal report assessing awareness of the target structures. Participants were reminded that the sentences of the language were formed by two structures and they were encouraged to verbalize any knowledge about them. More precisely, they were asked:

1. As mentioned during the experiment, the sentences you listened to were formed according to two structures. Did you notice which structures were these?
2. If yes, please indicate which you think these structures are.

These questions appeared on the screen, with a text box below in which participants typed their answers. Since learning would occur under incidental conditions, it was expected to be implicit and, hence, to produce unconscious, non-verbalizable knowledge of the similar and the dissimilar structures (cf. Chapter 1, Section 1.3). Yet, if learners had conscious knowledge of the structures, they were expected to provide metalinguistic descriptions of them, for instance:

- **Similar structure:** A structure contains the word *que* and then a verb conjugated (in the present subjunctive).
- **Dissimilar structure:** A structure does not contain the word *que* and contains a verb in the infinitive.

2.5. Predictions

In Section 2.1, I presented my hypotheses regarding the facilitative role of cross-linguistic syntactic similarity in learning the similar structure compared to the dissimilar one. In brief, I assumed that the similar structure would be processed using an L1 structure present in the linguistic system at the initial state of L2 acquisition, but that the dissimilar structure would need to be acquired as a result of exposure. In this light, I hypothesized that after exposure to the two structures there could be two scenarios, depending on whether this was or was not enough to trigger learning of the dissimilar structure. On the one hand, **Hypothesis 1 (H1)** claimed that the similar structure would be established in learners' linguistic system, but the dissimilar structure would be not. On the other hand, **Hypothesis 2 (H2)** claimed that the similar and the dissimilar structure would both be established in learners' linguistic system, but that the former would be more robustly established than the latter. Whether one of these hypotheses was met or not will be seen in the testing phase. As will be detailed in the next section, accuracy and *d'* analyses assessed learners' performance in the GJT. Learning of a structure would be reflected by learners' ability to accept grammatical sentences formed by that structure and reject ungrammatical sentences violating it (accuracy analysis) or by evidence that learners were sensitive to the difference between grammatical sentences formed by a structure and their ungrammatical counterparts (*d'* analysis). This is consistent with Chomsky's (1965) observation that adults, who know how the grammar of their native language works, can tell apart grammatical sentences from ungrammatical ones as a consequence of having *linguistic competence*. The same reasoning is applied to L2 acquisition.

If H1 was correct, I predicted that learners would judge the grammatical similar structure and the ungrammatical similar one significantly above chance, but this would not occur for the grammatical and the ungrammatical dissimilar structures. Additionally or alternatively, learners would be sensitive to the difference between the grammatical similar structure and its ungrammatical counterpart, but not to the difference between the grammatical dissimilar structure and its ungrammatical counterpart. Since I assumed that the similar structure would be firmly established in learners' minds from the start of L2 acquisition, I also predicted that accuracy for the grammatical and the ungrammatical similar structures and/or sensitivity to the difference between the two would be largely above chance. This would reflect a good ability to discriminate between the two constructions.

By contrast, **if H2 was correct,** I predicted that learners would judge the two grammatical structures and the two ungrammatical structures significantly above chance, and that accuracy would be significantly higher for the similar structure and its ungrammatical counterpart than for the dissimilar structure and its ungrammatical counterpart. Additionally or alternatively, learners would be sensitive to the difference between the similar/dissimilar structures and their ungrammatical counterparts, and sensitivity to the grammatical vs. the ungrammatical similar structure would be significantly higher than sensitivity to the grammatical vs. the ungrammatical dissimilar structure.

2.6. Coding and data analysis

I used the programming environment R (R Core Team, 2022, version 4.2.2) to analyse the data of the experiment. The function and package used to compute each statistical test and effect size measure are only stated the first time that a test or effect size is mentioned.

2.6.1. Exposure phase

I measured the accuracy of plausibility judgements in the four exposure conditions (plausible similar structure, implausible similar structure, plausible dissimilar structure and implausible dissimilar structure). Responses in each trial of the plausibility judgement task were coded as binary (1 = correct plausibility judgement, 0 = incorrect plausibility judgement). Two trials, displaying a plausible dissimilar structure and an implausible dissimilar structure, were excluded from the analysis because participants could not provide a plausibility judgement in the time allowed (5 seconds). E-prime automatically coded accuracy in those trials as 0 when, actually, no response was given. First, accuracy was analysed descriptively. Then, I analysed the effect of sentence plausibility and cross-linguistic syntactic similarity on accuracy. Because accuracy is a binary variable, I used a generalized linear mixed effects model fitted with the function *glmer* from the *lme4* package (D. Bates, Mächler, et al., 2015). I used deviation contrast coding for the variables Plausibility and Cross-linguistic similarity, assigning the value 0.5 to the categories *Plausible* and *Similar structure* and the value -0.5 to the categories *Implausible* and *Dissimilar structure.* The model fitted had Accuracy as a dependent variable, the interaction between Plausibility and Cross-linguistic similarity as a fixed effect and random intercepts by participant and by item. By-participant random slopes of Plausibility and Cross-linguistic similarity were included to account for the fact that the effect of these variables on accuracy could be different for each participant. Following D. Bates, Kliegl et al. (2015), random slopes were only included if they improved the model's likelihood, as assessed by nested model comparisons[10].

2.6.2. Testing phase

Participants' responses in the GJT were coded as binary (1 = correct grammaticality judgement, 0 = incorrect grammaticality judgement). Trials in which participants could not provide a judgement in the time allowed (5 seconds) were excluded from the analysis, since, as mentioned, E-prime coded accuracy in these trials as 0 when actually no judgement was made. Overall, 1.09% (21/1920) of trials were excluded. This corresponded to 1.46% (7/480) of grammatical similar structures, 1.88% (9/480) of grammatical dissimilar structures, 0.25% (1/480) of ungrammatical similar structures and 0.83% (4/480) of ungrammatical dissimilar structures.

Accuracy and d' analyses

Performance in the GJT was analysed by looking at accuracy percentages and *d'* scores, the latter taken from the Signal Detection Theory (Green & Swets, 1966). The difference between accuracy and *d'* analyses is that the significance tests used to analyse accuracy data in a GJT (e.g. t-tests and mixed effects models) cannot account for response bias —the tendency to accept or reject sentences irrespective of their grammaticality. By contrast, *d'* is an index of sensitivity to the difference between grammatical and ungrammatical sentences unaffected by response bias (Y. Huang & Ferreira, 2020)[11]. In what follows, I introduce the main aspects of the Signal Detection Theory, including how the *d'* analysis works and how it may be interpreted (see Macmillan & Creelman, 2005; Stanislaw & Todorov, 1999 for more details). The Signal Detection Theory was initially applied in perception studies to assess discrimination between *signals* or stimuli and *noise* or no stimuli. In the last decades, it has been used to

[10] Deviation coding, also known as sum coding, allows to interpret the effect of the independent variables on the dependent variable as main, "omnibus" effects when interactions are present (Sonderegger et al., 2018).
In the present analysis, a by-participant random slope of Plausibility*Cross-linguistic similarity did not improve the model's likelihood ($X^2(4) = 1.49$, $p = .83$).

[11] Y. Huang and Ferreira discuss acceptability judgements instead of grammaticality judgements, but their reasoning is valid for both types of tasks.

analyse sensitivity in tasks such as the GJT, during which participants are presented with *old* or previously heard items (grammatical sentences) and *new* or previously unheard items (ungrammatical sentences) and have to indicate whether an old or a new stimulus was presented. To do this, participants are assumed to rely on a *decision variable*, to which they attribute a value in each trial (in Experiment 1, how familiar each structure is, compared to the ones presented in the exposure phase). The value of the decision variable is compared against a *selection criterion,* which is the value that the decision variable has to achieve for a stimulus to be judged as old (accepted) or new (rejected). The distance between the selection criterion and the middle point between the signal and noise distributions is the *index of bias c* (criterion location), indicating whether participants are biased towards accepting or rejecting the stimuli.

In a GJT, learners can correctly identify a grammatical sentence as grammatical (*Hit*) or they can incorrectly judge it as ungrammatical (*Miss*). Likewise, they can correctly identify an ungrammatical sentence as ungrammatical (*Correct Rejection*) or they can judge it as grammatical (*False Alarm*). Considering this, *d′* scores compare the probability of identifying a grammatical sentence as such (hit rate) with the probability of misjudging an ungrammatical sentence as grammatical (false alarm rate). In Experiment 1, the hit rate corresponds to the probability of judging a grammatical sentence as "correct", while the false alarm rate is the probability of misjudging an ungrammatical sentence as "correct". Since the hit rate and the false alarm rate for each participant may come from distributions with different means and standard deviations, comparing raw values might not always be adequate. To make the two rates comparable, the inverse normal transform or *z-transform* is applied to the data. With this operation, probabilities are re-expressed as standard or *z scores* which come from a standard normal distribution and express how many standard deviations each participant's hit rate and false alarm rate are away from the mean. The z-transform equals infinity when the hit rate or the false alarm rate is 0 or 1, so the former is adjusted to .01 and the latter, to .99. To get the *d′* or sensitivity measure to the difference between signal and noise, the z-transform of the false alarm rate is subtracted from the z-transform of the hit rate. A *d′* score of 0 indicates incapacity to discriminate between grammatical and ungrammatical stimuli (chance performance). Perfect discrimination corresponds to a *d′* score of 4.65. Negative *d′* scores, up to -4.65, are also possible and reflect that the false alarm rate is higher than the hit rate. The *d′* score is unaffected by response bias because bias is assumed to affect the hit rate and the false alarm rate equally. The index of bias *c* is calculated as the negative value of half of the sum of the z-transforms of the hit rate and the false alarm rate. A *c* value of 0 indicates no bias to accept or reject the stimuli. A value significantly higher than 0 reflects a preference for rejecting most stimuli, whereas a value significantly lower than 0 indicates a preference for accepting them.

The similar and the dissimilar structure as part of the linguistic system

To test the predictions in Section 2.5, I first had to assess whether the similar and the dissimilar structure were part of learners' linguistic system. Starting with the accuracy analysis, I first examined the mean accuracy percentages when judging the grammatical and ungrammatical similar and dissimilar structures, both descriptively and visually. All graphs in this chapter were generated with the function *ggplot* from the *ggplot2* package (Wickham, 2016). Then, one sample t-tests fitted with the function *t.test* from the *stats* package (R Core Team, 2022) compared mean accuracy percentages against chance (50%). *Cohen's d* was calculated as a standardised measure of effect size for all t-tests using *cohens_d* from the *rstatix* package (Kassambara, 2021). Following Cohen (1988), I considered a *d* of 0.2, 0.5 and 0.8 to be small, medium and large, respectively[12]. In addition, to have a better understanding of the grammaticality judgements for each structure, I calculated mean accuracy percentages in the four experimental conditions for all participants individually. As will be detailed in Section 2.7.2, overall learners could not judge nor grammatical sentences formed by the dissimilar structure nor ungrammatical sentences violating the similar or the dissimilar structure above chance. I calculated whether accuracy in the two ungrammatical conditions significantly differed using a generalized linear mixed effects model. The model had Accuracy as a dependent variable, Cross-linguistic similarity as an independent variable, random intercepts by participant and by item and a random slope of Cross-linguistic similarity by participant. Treatment coding was used for the variable Cross-linguistic similarity, with the category *Similar structure* coded as 0 and the category *Dissimilar structure,* as 1.

Next, I examined whether participants had a tendency to accept or reject sentences in the GJT. I coded responses in the test as Hits, False alarms, Misses or Correct rejections. Then, I calculated the index *c* examining response bias when judging grammatical and ungrammatical sentences, generally, and grammatical and ungrammatical sentences following or violating the similar structure, particularly. I calculated these indices for each participant individually using the function *dprime* from the *psycho* package (Makowski, 2018). Later, one-sample t-tests determined if the mean indices of response bias were significantly above or below chance (i.e. zero). Since the tests yielded a significant response bias, I performed a *d'* analysis. I used the function *dprime* to calculate participants' *d'* scores reflecting sensitivity to the difference between the grammatical similar structure and the ungrammatical similar structure, on the one hand, and the grammatical dissimilar structure and the ungrammatical dissimilar structure, on the other hand. One-sample t-tests examined if mean *d'* scores were significantly different from zero. As will be shown in Section 2.7.2, the results of the *d'* analysis suggested that Hypothesis 1 could be retained, although not conclusively, so no further analyses were performed.

2.6.3. Debriefing phase

Participants' responses in the verbal report were transcribed. Two experimenters used a rubric to classify participants as *aware* or *unaware* of the similar and the dissimilar structure (the transcriptions and the rubric can be consulted in Appendix A-8). In short, *aware*

[12] This is the effect size reported for all t-tests in this chapter. Unless otherwise specified, it must be assumed that *Cohen's d* was calculated after each t-test.

participants were those who could provide a metalinguistic description of one or the two target structures, as exemplified in Section 2.4.3. *Unaware* participants were those who were not able to identify the varying elements in the structures (the presence/absence of the complementizer *que* and the finite or non-finite nature of the embedded verb) or, alternatively, were able to identify them but could not appropriately correlate them. Disagreements were discussed until a unanimous decision was made. Awareness results were coded as binary (1 = aware participant, 0 = unaware participant) and were analysed descriptively.

2.7. Results

2.7.1. Exposure phase

Table 2.4 summarises mean accuracy percentages for the plausible similar structure (Pl.SS), the implausible similar structure (Impl.SS), the plausible dissimilar structure (Pl.DS) and the implausible dissimilar structure (Impl.DS). In all cases, percentages are above 90%, which suggests that participants understood the sentences in the plausibility judgement task very well. A generalized linear mixed effects model indicated that accuracy was not significantly affected by Plausibility (β = -0.33, *SE* = 0.44, *z* = -0.76, *p* = .45), Cross-linguistic similarity (β = -0.13, *SE* = 0.35, *z* = -0.38, *p* = .70) or the interaction between the two variables (β = 0.46, *SE* = 0.63, *z* = 0.73, *p* = .47).

	Pl.SS	**Impl.SS**	**Pl.DS**	**Impl.DS**
M	95.48	93.69	93.82	93.32
SD	20.78	24.94	24.34	24.98
95%CI	[93.82, 97.15]	[91.74, 95.64]	[91.89, 95.76]	[91.32, 95.33]

TABLE 2.4. Mean accuracy (%), standard deviations (%) and 95% confidence intervals for all conditions in the PJT of Experiment 1. *M* = Mean, *SD* = Standard Deviation, *95%CI* = 95% Confidence Interval.

2.7.2. Testing phase

Table 2.5 and Figure 2.2 summarise mean accuracy percentages for the grammatical similar structure (SS), the grammatical dissimilar structure (DS) and their ungrammatical counterparts (*SS and *DS). As shown, learners accepted a mean of 73.57% of SS items (significance from chance, *t* (23) = 4.35, *p* < .001, *d* = 0.89) and a by-participant analysis revealed that 83.33% (20/24) of learners accepted more than 50% of these sentences. By contrast, mean accuracy for DS items was at chance (49.89%, *t* (23) = -0.04, *p* = .97, *d* = -0.01). A look at learners' individual performance revealed that just 17% (4/24) performed at chance; 33% (8/24) accepted more than 50% of DS sentences and the remaining 50% of learners (12/24) rejected more than 50% of DS items. Turning to the ungrammatical conditions, overall learners rejected a mean of 37.79% of *SS items and accuracy was significantly below chance (*t* (23) = -2.21, *p* = .02, *d* = -0.45). Individually, just 29.17% (7/24) of learners rejected more than 50% of *SS sentences. Similarly, learners rejected a mean of 41.18% of all *DS items and this percentage was significantly below chance (t (23) = -1.45, p = .04, d = -0.30). Individually, just 20.83% (5/24) of learners rejected more than 50% of *DS sentences. Descriptively, learners rejected a higher percentage of *DS items than *SS items. Yet, accuracy did not significantly differ between the two conditions (β = 0.16, SE = 0.74, z = 0.22, p = .83).

Overall, these results provide no evidence that the similar structure or the dissimilar structure were part of learners' grammar, for learners could not distinguish grammatical sentences formed by these structures from their ungrammatical counterparts. Thus, the accuracy analysis did not support nor Hypothesis 1 nor Hypothesis 2.

	SS	***SS**	**DS**	***DS**
M	73.57***	37.79*	49.89	41.18*
SD	44.14	48.54	50.05	49.27
95%CI	[69.58, 77.56]	[33.43, 42.14]	[45.36, 54.43]	[36.74, 45.61]

TABLE 2.5. Mean accuracy (%), standard deviations (%) and 95% confidence intervals for all conditions in the GJT of Experiment 1. *M* = Mean, *SD* = Standard Deviation, *95%CI* = 95% Confidence Interval. Significance from chance: * p < .05.*** p < .001[13].

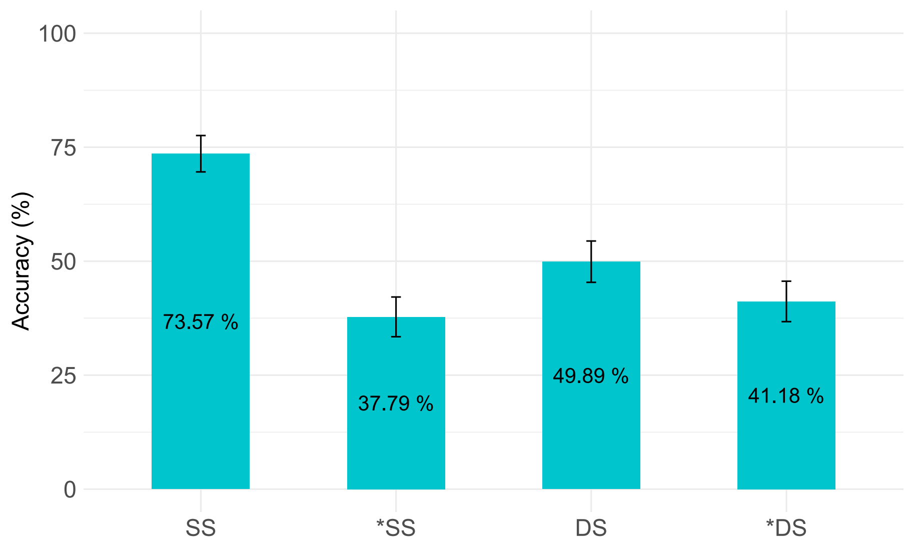


FIGURE 2.2. Mean accuracy (%) across conditions in the GJT of Experiment 1. Error bars represent 95% confidence intervals.

[13] T-tests comparing mean accuracy percentages for SS, *SS and *DS against chance were one-tailed, for I had a directional alternative hypothesis, i.e. accuracy was significantly greater or lower than 50%. The t-test comparing the mean accuracy percentage for DS against chance was two-tailed, for I had a non-directional alternative hypothesis, i.e. accuracy was statistically different from 50% (Levshina, 2015).

The previous analysis suggests that learners tended to accept most sentences, generally, and most SS and *SS sentences, particularly. This was confirmed by the index *c* indicating response bias when judging grammatical and ungrammatical sentences and when judging SS and *SS items. In both cases, the mean index was negative and significantly below zero (M = -0.30, SD = 0.35; t (23) = -4.12, p <.001, d = -0.84 and M = -0.56, SD = 0.84; t (23) = -3.20, p = .002, d = -0.65, respectively). As mentioned in Section 2.6.2, accuracy does not separate sensitivity to the difference between grammatical and ungrammatical sentences from response bias. Therefore, learners could be sensitive to the difference between the grammatical similar structure and the ungrammatical similar structure, but this could be masked in the accuracy analysis due to the tendency to accept these structures. Table 2.6 and Figure 2.3 present the *d’* scores reflecting sensitivity to the difference between SS and *SS items and between DS and *DS items separated from response bias.

	SS vs. *SS	**DS vs. *DS**
M	0.43**	-0.21
SD	0.59	0.52
95%CI	[0.39, 0.46]	[-0.24, -0.18]

TABLE 2.6. Mean *d’* scores, standard deviations and 95% confidence intervals reflecting sensitivity to SS vs.*SS and DS vs.*DS in the GJT of Experiment 1. *M* = Mean, *SD* = Standard Deviation, *95%CI* = 95% Confidence Interval. Significance from zero: **p <.01[14].

[14] T-tests comparing *d’* scores against chance were one-tailed for SS vs. *SS items, for I tested whether sensitivity was significantly greater than zero, and two-tailed for DS vs. *DS items, for I tested whether sensitivity was statistically different from zero.

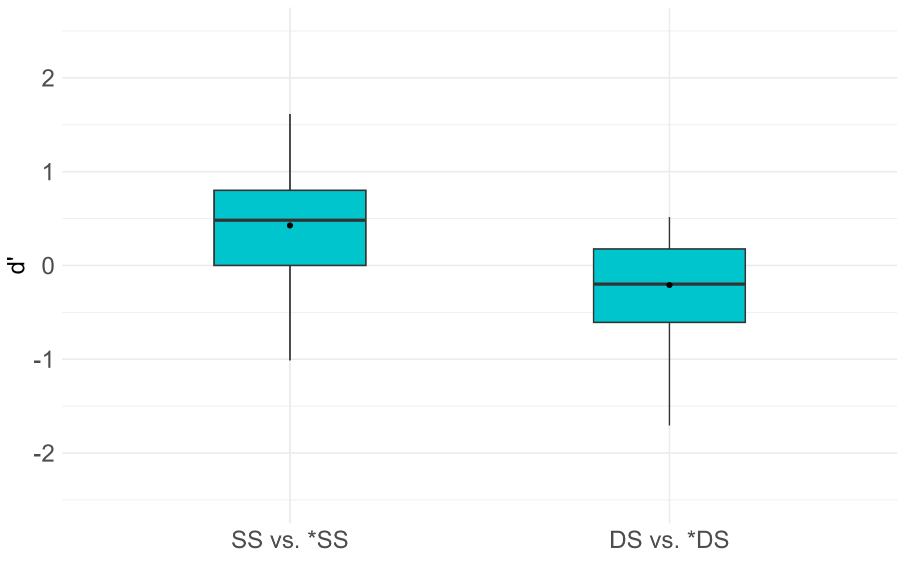


FIGURE 2.3. Distribution of *d'* scores reflecting sensitivity to SS vs.*SS and DS vs.*DS in the GJT of Experiment 1. The black horizontal line in the box of the boxplot shows the median. The black dot is the mean. Whiskers end at the smallest and largest data points that fall within 1.5 times the interquartile range from the first quartile (25%) and the third quartile (75%) of the data.

As shown, the mean *d'* score reflecting sensitivity to DS vs. *DS items was negative, which shows that the number of times that *DS items were misjudged as grammatical or "correct" was larger than the number of times that DS items were identified as such. This *d'* score was not statistically different from zero (at most, it was marginally below zero; t (23) = -1.87, p = .07, d = -0.38). This indicates, in line with the accuracy analysis, that learners were not able to discriminate between the dissimilar structure and its ungrammatical counterpart[15]. Interestingly, the mean *d'* score reflecting sensitivity to SS vs. *SS items was significantly above zero (t (23) = 3.44, p = .001, d = 0.70). Contrary to the accuracy analysis, this shows that learners were able to discriminate between the grammatical similar structure and the ungrammatical similar one. Yet, this result should be interpreted with caution. As advanced in Section 2.6.2, a *d'* score of zero reflects incapacity to discriminate between grammatical and ungrammatical stimuli (0% sensitivity). The maximum *d'* score is 4.65 and reveals perfect discrimination (100% sensitivity). The mean *d'* score showing sensitivity to the difference between the grammatical and the ungrammatical similar structure was 0.43, which is 90.75% away from perfect sensitivity and, thus, corresponds to a very poor discrimination ability. In fact, the hit rate for SS items was 0.73, indicating that learners correctly identified them as grammatical on most occasions. However, the false alarm rate for *SS items was 0.62,

[15] When *d'* is negative, the parametric index of sensitivity might be affected by response bias (Rogers, Révész, et al., 2015). In those cases, the non-parametric index of sensitivity *A'* can be calculated. The value of *A'* ranges between 0 and 1, with a value of 0.5 indicating no sensitivity. The mean *A'* score reflecting sensitivity to the difference between DS and *DS items was (0.45, *SD* = 0.17, 95%CI = [0.43, 0.46]) and it was not significantly different from 0.5 (t (21) = -1.39, p = .18, d = -0.30). Thus, it corroborated the result of the parametric test.

revealing that learners frequently misjudged these sentences as grammatical, as already mentioned in the accuracy analysis.

In sum, the *d*' analysis suggests that Hypothesis 1, stating that the similar structure would be established in learners' linguistic system but the dissimilar structure would be not, might be retained. I predicted that if this hypothesis was correct, learners would be sensitive to the difference between the grammatical and the ungrammatical similar structures, but not to the difference between the grammatical and the ungrammatical dissimilar structures, as observed. Nevertheless, I also predicted that, since the similar structure would be a very stable item in learners' linguistic system already when L2 acquisition started, sensitivity to the difference between SS and *SS would be quite large, indicating a good ability to discriminate between the two. This prediction was not met.

2.7.3. Debriefing phase

All participants were coded as *unaware* of the similar and the dissimilar structure according to their verbal reports, with the exception of one participant (P1) who provided a correct description of the similar structure[16]. Within the 23 unaware participants, 13 did not answer to the verbal report. Eight participants commented that a structure contained the word *que* and the other did not. Within these, one (P24) additionally mentioned that the ending of the (embedded) verb varied, but could not correlate the presence/absence of *que* with the finite or non-finite verb. The remaining two unaware participants did not comment nor on the presence/absence of *que* nor on the finite/ non-finite nature of the embedded verb. Specifically, P8 reported that the sentences in the experiment always started with a phrase of the type *it is important* or *it is necessary* and then contained a verb. P18 incorrectly mentioned that the proper noun appeared in different positions in the sentence (beginning-middle-end).

2.8. Discussion

In Experiment 1, I investigated the facilitative role of cross-linguistic syntactic similarity at the initial stage of adult L2 syntax acquisition. Specifically, I explored whether Spanish natives with no knowledge of Galician showed a learning advantage for a Spanish-Galician similar structure, existing in the L1 and the L2, over a Spanish-Galician dissimilar structure, expressing the same meaning but existing only in the L2. I incidentally exposed learners to the structures in a Plausibility Judgement Task (PJT). Then, I tested learning of the structures in a Grammaticality Judgement Task (GJT). The experiment ended with a verbal report. Based on the MOGUL framework (Sharwood Smith & Truscott, 2014) and several theories and models of L2 acquisition such as the Full transfer/Full access model (FT/FA model, B. D. Schwartz & Sprouse, 1994, 1996), the Autonomous Induction Theory (AIT, Carroll, 1999, 2001), the Unified Competition Model (UCM, MacWhinney, 2005) and the Micro-cue model of L2 acquisition (Westergaard, 2021), I hypothesized that learners would process the cross-linguistically

[16] P = Participant

similar L2 structure using the equivalent L1 structure in their linguistic system. Conversely, they would need to acquire the cross-linguistically dissimilar L2 structure during exposure. The results of the GJT suggested that the experiment was not suited neither to elicit knowledge of the similar structure nor to prompt learning of the dissimilar structure. On the one hand, the test provided only weak evidence that the similar structure was part of learners' linguistic system. That is, learners mostly accepted the similar structure but could not reject its ungrammatical counterpart and, while there was some sensitivity to the difference between the two, this was smaller than expected for a structure firmly established in learners' minds at the beginning of L2 acquisition. On the other hand, there was no evidence that the dissimilar structure was learnt. Learners could not judge neither sentences formed by this structure nor ungrammatical sentences violating it significantly above chance, and they were not sensitive to the difference between the two. In the next sections, I discuss the results and possible shortcomings of the exposure and the testing phases, as well as the results of the debriefing phase. I conclude by discussing the challenge of conducting implicit learning experiments.

2.8.1. Discussion of the exposure phase

Plausibility judgments were at ceiling for plausible and implausible sentences formed by the similar or the dissimilar structure, which suggests that participants comprehended very well the proposition expressed by all sentence types. This matches previous implicit learning experiments using a PJT to expose novice learners to L2 structures (e.g. Rebuschat, 2009; Williams & Kuribara, 2008). Nevertheless, many experiments using this task do not report its results (e.g. Kim & Fenn, 2020; Maie & DeKeyser, 2020; Rebuschat & Williams, 2012; Tagarelli et al., 2016). In Experiment 1, no condition seemed to be more challenging to understand than the rest, since accuracy did not significantly differ as a function of Plausibility, Cross-linguistic similarity or their interaction. In addition, all sentences were made up of Spanish-Galician cognates. The fact that in spite of not knowing any Galician, learners understood sentences very well suggests that, as hypothesized in Section 2.1, cognates were processed as if they were equivalent to their L1 translations. I propose that this could be explained within MOGUL. In short, following Jackendoff (2002), MOGUL defines a word as a chain of phonological, syntactic and conceptual representations (Sharwood Smith & Truscott, 2014, sec. 11.6.4). When learners hear a word, its phonological representation activates in the phonological module, as well as those phonological representations sharing features with it (Sharwood Smith & Truscott, 2014, sec. 3.3.5). In the case of a cognate, this would include the phonological representation of its L1 translation, which, in turn, would trigger the activation of its syntactic and conceptual representations in adjacent modules. To process a word, its phonological representation needs to be assigned a syntactic and a conceptual representation (Sharwood Smith & Truscott, 2014, sec. 4.2). For the L2 cognate, these would be the syntactic and conceptual representations of its L1 counterpart.

In Experiment 1, processing and/or acquisition of the similar and the dissimilar structure should have occurred during the PJT. However, in spite of learners' good understanding of exposure sentences, the testing phase provided weak or no evidence that the similar and the dissimilar structure were learnt. The PJT is used in implicit learning experiments, where learners are expected to process and accommodate syntactic structures or features while their attention is on sentential meaning, not form (see Chapter 1, Section 1.3). In line with this, Sharwood Smith and Truscott maintain that for processing and learning to occur, learners do not need to pay attention to the form of the input. More precisely, for a structure to be processed and/or acquired, an auditory or visual representation of the sentence including that structure must be created. This corresponds to the so-called *global awareness of the input.* Then, a representation of the target structure must follow. This process, referred to as *noticing,* is claimed to be equivalent to the concept in Schmidt's *Noticing Hypothesis* (1990, 1992). This follow-up representation becomes *intake* and is fed to the phonological and syntactic processors (Sharwood Smith & Truscott, 2014, sec. 8.4.3, 9.3.2). The distinction between *input* and *intake* was first postulated by Corder (1967) and, since then, these concepts have been discussed and incorporated in many theories and models of second language acquisition (e.g. Carroll, 1999; Chaudron, 1985; Gass, 1997; VanPatten, 1996, 2002, 2004; VanPatten & Cadierno, 1993). I propose that two characteristics of the exposure phase might have prevented either the similar and the dissimilar structure, or just the dissimilar structure, from becoming intake. In addition, a third characteristic could have caused that the dissimilar structure became intake, but that it did not become part of learners' linguistic system.

First, the plausibility of exposure sentences was determined by the combination of the embedded transitive verb and its direct object (e.g. *É importante que Pedro* <u>*repare a radio*</u>*,* "it is important that Pedro fixes the radio" (plausible sentence) vs. *É importante que Pedro* <u>*repare o tornado*</u>*,* "it is important that Pedro fixes the tornado" (implausible sentence)). This means that plausibility judgements could be made by paying attention only to the embedded verb and the following noun phrase in each sentence, i.e. avoiding processing the entire sentences, which were structurally quite complex due to the presence of the embedded clause. Thus, it is possible that, during the exposure phase, learners' attention was mostly focused on the verb-object combination in each sentence. This could have prevented learners from constructing an auditory representation of the similar and the dissimilar structure that could be subsequently processed and/or acquired. In short, it could be that the two structures were in the input, but did not become intake.

Second, even if quite more unlikely, it could be that the fact that exposure sentences were presented in auditory form hampered a correct processing of the dissimilar structure, preventing it from being established in learners' linguistic system. To understand why this is a possibility, it must be noted that, in Spanish (participants' L1), finite verbs end in a vowel and infinitives end in *–ar, –er* or *–ir.* One of the forms in which Spanish can express the past tense of verbs ending in *–ar* in the subjunctive mood coincides with the combination of the infinitive + *a* (e.g. *firmara la carta,* "signed$_{\text{SBJV}}$ the letter" vs. *firmar la carta,* "to sign the letter"). In Experiment 1, the infinitive verb in the dissimilar structure (also ending in *–ar, –er* or *–ir*) was followed by a determiner which took the form of *a* ("the", feminine) or *o* ("the", masculine).

The fact that participants could only listen to the sentences (and not read them), could have caused that the infinitive verb was processed as a finite verb ending in a vowel by segmenting the verb and the following determiner as a unique word. Simply put, the presence of an infinitive verb in the input could have been irrelevant for syntax learning, i.e. the auditory representation of the dissimilar structure fed to the syntax processor could have not contained a non-finite form. In this case, only the similar structure would have become intake.

Finally, a third possibility is that the similar and the dissimilar structure became intake but that, as hypothesized (Hypothesis 1, H1), the number of times that the dissimilar structure was processed during exposure was not enough for it to become part of learners' linguistic system. Learners were first exposed to the dissimilar structure in Experiment 1. In MOGUL terms, this means that this structure had no pre-existing resting activation level. Each time that the structure was processed, its resting activation level increased, so that it gradually became a stable item in learners' linguistic system (Sharwood Smith & Truscott, 2014, sec. 4.2). It could be that by the end of the exposure phase, the resting activation level of this structure was still too low for it to be part of L2 grammar. Conversely, the number of instances of the similar structure should not have been problematic, since it was already part of learners' linguistic system prior to the exposure phase and it had a high resting activation level due to its past use in the L1 (Sharwood Smith & Truscott, 2014, sec. 10.3). The difficulty learning the dissimilar structure could also be explained by approaches to L2 acquisition other than MOGUL. For instance, in terms of B.D. Schwartz and Sprouse's FT/FA model and Carroll's AIT, it could be that this structure was not learnt because input was not enough to allow the grammar to be restructured, either by accessing UG or by using a learning mechanism, respectively. According to MacWhinney's UCM, input would not have been sufficient for learners to acquire the L2 cue corresponding to the dissimilar structure. Finally, in terms of Westergaard's Micro-cue model of L2 acquisition, learners would not have been able to acquire the micro-cue consistent with the dissimilar structure.

In sum, I propose that at the end of the exposure phase the dissimilar structure was not established in learners' linguistic system, which would explain why performance in the testing phase did not reveal learning of this structure. However, I also propose that the similar structure was firmly established in learners' linguistic system at the beginning of L2 acquisition. If this is true, why was this not conclusively reflected in the GJT? In addition, how can one account for the specific results of the accuracy and *d'* analyses assessing grammaticality judgements for the grammatical and the ungrammatical similar and dissimilar structures? I address these questions in the next section.

2.8.2. Discussion of the testing phase

Following the exposure phase, participants were informed that they had listened to sentences formed according to two structures, and that they would listen to new sentences and would need to decide, as quickly as possible, whether these followed the structures of the language or not. Participants were instructed to use their intuition to perform the task, a behaviour that was further encouraged by time pressure. Hence, they were not told to reflect about the form of the sentences or about their grammaticality judgements. Yet, I hypothesize that if participants had merely processed the sentences using the structures in their linguistic system, accepting those sentences that could be processed using these structures and rejecting the rest, they would have accepted the similar structure and rejected the other structures, a behaviour that was not observed. In addition, since I assumed that the similar structure was a very stable item in learners' linguistic system, sensitivity to the difference between the grammatical and the ungrammatical similar structure would have been quite high (probably higher than around 10%, the result in Experiment 1). The fact that these results were not obtained suggests that learners consciously decided to accept or reject some structures, in spite of time pressure and the instruction to use their intuition. This is consistent with the observation that grammaticality judgement tasks are likely to involve conscious reflection about the grammaticality of the stimuli presented (DeKeyser, 2003; Suzuki & DeKeyser, 2017; Vafaee et al., 2017). The fact that learners were told that there were *two* correct structures in the language probably prompted this behaviour too. In what follows, I present my hypotheses regarding how the grammatical and the ungrammatical similar and dissimilar structures were processed in the GJT and I propose explanations for the results observed.

I start with the grammatical similar structure (e.g. *É importante que Pedro repare a radio,* "it is important that Pedro fixes the radio"). I hypothesize that learners processed this structure using the construction acquired as part of the L1. The use of the L1 structure is even more likely if we consider that the sentences formed by the similar structure that learners processed were made up of Spanish-Galician cognates. That is, the activation of the Spanish (L1) counterparts of the Galician (L2) words could have strengthened the activation of the L1 syntactic representation, making it even more available for processing (see, e.g., Hartsuiker and Pickering, 2008 and Hopp, 2017 for related ideas on how cross-linguistic lexical activation via cognates may contribute to cross-linguistic syntactic activation). Because this L1 structure was firmly established in learners' linguistic system, they tended to accept it; overall learners judged over 70% of SS items as "correct" in the L2 and, individually, 83.33% of learners accepted more than 50% of all SS sentences. However, participants also rejected a mean of 26.43% of these items (and, individually, 16.67% of learners rejected more than 50% of these constructions). This could be attributed to the fact that, as advanced in the previous section, for some learners the similar structure might have not become intake during the exposure phase. These learners could have considered that, despite being present in the L1, the similar structure was not part of the L2 and, thus, rejected it.

Turning to the grammatical dissimilar structure (e.g. *É importante Pedro reparar a radio,* "it is important that Pedro fixes the radio"), results showed that mean accuracy when judging this structure was statistically indistinguishable from 50%. Yet, only four learners performed at chance. The rest either accepted this structure or rejected it. On the one hand, if the dissimilar structure did not become intake during exposure, this structure would have been inexistent in learners' linguistic system. This could have caused that, when processing it in the GJT, learners identified the mismatch between this structure and any structures learnt, and either rejected it as part of the L2 or performed at chance. On the other hand, it could be that the dissimilar structure became intake during exposure, but that the number of times that it was processed was not high enough for it to be a stable item in the linguistic system. In MOGUL terms, the structure would have had a very low resting activation level. If this was the case, then the dissimilar structure would have activated during processing in the GJT, but very weakly, since the starting point for the activation that an item receives during processing is its resting activation level (Sharwood Smith & Truscott, 2014, sec. 3.3.5). This weak activation could have caused learners to accept it, but also to reject it or perform at chance.

I hypothesize that when participants listened to the ungrammatical similar structure (e.g. **É importante que Pedro reparar a radio*), they processed it as if it was its grammatical counterpart up to the moment the embedded verb was perceived, when the ungrammaticality of the structure should have been detected. Detecting such ungrammaticality was probably the reason why around 38% of *SS sentences were judged to be "incorrect" in the L2, why 29.17% of learners rejected more than 50% of these sentences and why learners showed almost 10% sensitivity to the difference between the grammatical and the ungrammatical similar structure. However, as mentioned, this sensitivity was lower than expected, considering that the similar structure was firmly established in learners' linguistic system. Additionally, a mean of over 60% of *SS sentences were accepted and, individually, 70.83% of learners accepted 50% or more than all *SS items. An explanation for this could be the similarity between the ungrammatical similar structure and its grammatical counterpart (they both contained an embedded clause introduced by *que* ("that") and only differed in the finite or non-finite embedded verb). That is, only the similar structure (and not the dissimilar structure) was established in learners' linguistic system, but participants were told that there were *two* correct structures in the L2. In this light, the similarity between the grammatical and the ungrammatical similar structures could have led learners to consciously accept the ungrammatical construction, i.e. to consider it "correct" in the language.

Last, I must discuss grammaticality judgements for the ungrammatical dissimilar structure (e.g. **É importante Pedro repare a radio*). Irrespective of whether the (grammatical) dissimilar structure became intake during the exposure phase or not, the mismatch between the ungrammatical dissimilar structure and the structures in the linguistic system should have led participants to reject it in the GJT. This was the case on 41% of the occasions and, individually, 20.83% of learners rejected more than 50% of all *DS items[17]. Nevertheless, learners also

[17] One may be wondering whether an alternative explanation for the correct rejection of *DS sentences was that, for some learners, the dissimilar structure did become part of the grammar and, consequently, learners could identify the ungrammaticality of the ungrammatical dissimilar structure compared to the grammatical dissimilar one. This does not seem to be the case, for none of the learners who accepted more than 50% of grammatical dissimilar structures could reject more than 50% of their ungrammatical counterparts, nor was sensitive to the difference between the two structures.

accepted an average of almost 60% of *DS items and 79.17% of all learners accepted 50% or more than all *DS sentences. An explanation for this might also be the similarity between the grammatical similar structure and the ungrammatical dissimilar structure. That is, when processing the ungrammatical dissimilar structure, the finite embedded verb possibly weakly reactivated the grammatical similar structure. This is consistent with MOGUL's assumption that the activation of a syntactic representation (in this case, the representation of the verb) spreads to other syntactic representations containing it (in this case, the representation of the similar structure) (Sharwood Smith & Truscott, 2014, sec. 4.5.3). The fact that the ungrammatical dissimilar structure resembled the only one of the two L2 structures that was firmly established in learners' linguistic system could have led participants to accept it. This behaviour, together with the incapacity to accept sentences formed by the dissimilar structure significantly above chance, would explain why Experiment 1 found no sensitivity to the difference between the grammatical and the ungrammatical dissimilar structure.

Overall, the testing phase reveals a strong tendency to accept the grammatical similar structure (SS) and the two ungrammatical structures (*SS and *DS), which bear some similarity with the grammatical structure due to the presence of the complementizer (*SS) or the subjunctive embedded verb (*DS). Descriptively, the accuracy analysis indicated that participants tended to judge as "correct" more *SS items than *DS ones (even if the difference in accuracy between the two was not significant). In other words, participants tended to accept more ungrammatical embedded clauses introduced by a complementizer than ungrammatical embedded clauses not introduced by a complementizer. Considering the hypothesis that only the similar structure was established in learners' linguistic system, a possible explanation is that *SS was perceived as being more similar to SS than *DS was and, hence, participants considered more often that, if SS was correct in the L2, so was *SS.

2.8.3. Discussion of the debriefing phase

In the verbal report, no participant provided a description of the dissimilar structure and only one provided a description of the similar structure. The testing phase evidenced no knowledge of the dissimilar structure, so it is possible that learners did not report it simply because they had not learnt it. By contrast, the testing phase did capture some knowledge of the similar structure, so the verbal report suggests that any knowledge of this structure might have remained unconscious or non-verbalizable for all but one participant, matching the results of previous implicit learning experiments (e.g. Kim & Fenn, 2020; Leung & Williams, 2006; Rebuschat, 2009; Tagarelli et al., 2016; Williams, 2005). However, it must be taken into account that thirteen learners did not verbalize any knowledge in the verbal report. It is possible that these learners did not provide a description of the similar structure because they did not have any explicit knowledge of it, but it is also possible that they did have explicit knowledge of the structure but were not confident enough to report it. Alternatively, they could have simply decided not to respond to the task. That the inability to verbalize knowledge in the verbal report can be motivated by participants not being confident enough or simply choosing not to respond despite being aware of the learning target is a shortcoming of this task. Yet, I chose the verbal report as awareness measure because it is the most used in the literature and because, although several alternatives exist, such as confidence ratings and source attributions, they are not without their criticisms (see Rebuschat, 2013).

Within the eleven participants who did provide an answer to the verbal report, eight commented that the structures varied in the presence or absence of the word *que*. By contrast, just one commented that the ending of the (embedded) verb changed, which I interpret as referring to the finite vs. non-finite nature of the verb, a feature encoded in verbal morphology. A possible explanation for this difference resides in the characteristics of the complementizer and the infinitive and subjunctive verb endings in Spanish, participants' L1. The three elements are short and not stressed in connected speech. Yet, while the complementizer is a free form that stands alone as a word, verb endings are bound morphemes that cannot be pronounced separated from the verbal root (Real Academia Española & Asociación de Academias de la Lengua Española, 2009). In addition, infinitive and subjunctive forms are more frequent than the complementizer (infinitive, 29.05 occurrences per million; subjunctive, 7.97 occurrences per million; complementizer *que*, 4.36 occurrences per million; Corpus of the 21st Century Spanish, CORPES XXI, Real Academia Española, 2021). Stand-alone words are easier to perceive than bound morphemes and high frequency items are less noticeable than low frequency ones (Cintrón-Valentín & Ellis, 2016). All this suggests that the complementizer *que* may have been more salient than the infinitive and subjunctive verb endings and, thus, easier to report. An explanation based on saliency also goes in line with the finding that, in the testing phase, participants found it (non-significantly) easier to reject the ungrammatical embedded clause not introduced by *que* and containing a finite verb (*DS) than the ungrammatical embedded clause introduced by *que* but containing a non-finite verb (*SS). If as hypothesized, only the similar structure was part of learners' linguistic system, then this suggests that in ungrammatical sentences the absence of the complementizer was more salient than the non-finite embedded verb.

2.8.4. The challenge of conducting implicit learning experiments

Previous studies investigating initial L2 syntax learning under incidental conditions have acknowledged the difficulty of designing experimental paradigms that allow learning to take place and have suggested that coming up with such paradigms often requires more than one experiment. However, experiments that fail to find an initial learning effect are not usually reported in the literature, even if doing this could inform researchers about which experimental designs to avoid. As a matter of example, consider Rebuschat's (2009) doctoral dissertation and Rogers, Révész et al.'s (2015) set of experiments. Rebuschat conducted six experiments investigating learning of a language with English vocabulary and German syntax. In all experiments, the learning target were German verb-placement patterns or "rules", all different from English (e.g. "V2: Finite verb placed in second phrasal position of main clauses that are not preceded by a subordinate clause", p. 36). Participants were English natives with no knowledge of German. The procedure of the six experiments was similar. First, learners were exposed to the L2 verb-placement patterns. Then, learning of these patterns was tested in a grammaticality judgement task. Learning was defined as the ability to distinguish grammatical sentences formed by the L2 patterns from ungrammatical sentences violating these patterns. In Experiment 1, participants had to learn four verb-placement patterns, to which they were incidentally exposed by means of a plausibility judgement task with feedback. The results of the grammaticality judgement task showed no syntax learning; participants judged grammatical sentences above chance, but not ungrammatical ones. In Experiment 2, Rebuschat altered exposure sentences so that they displayed more prominently the position that the verb phrase occupied in the sentence. Yet, performance in the test mirrored that of Experiment 1. In Experiment 3, the patterns to be learnt were reduced from four to three and an elicited imitation task was added to the exposure phase. Nevertheless, syntax learning did not occur. In Experiment 4, the lexicon of the language was changed to nonsense words to remove any obstacles that lexical knowledge could pose to syntax learning. However, the verb-placement patterns were not learnt. In Experiment 5, learners were informed that the experimental sentences consisted of either a main clause or a main clause and a subordinate clause, which could appear in different orders. Nevertheless, this modification did not produce the learning desired. Finally, in Experiment 6, the exposure task changed from a plausibility judgement task to a rule-search task, which ultimately triggered learning of the L2 verb-placement patterns.

Turning to Rogers, Révész et al. (2015), they carried out three experiments examining whether English natives with no knowledge of Slavic languages learnt Czech case marking (nominative *–a*, accusative *–u*, instrumental *–ou*). In Experiment 1, participants were exposed to English sentences containing a Czech noun with one of the three case markers. Sentences could be formed according to four syntactic patterns. Participants listened to each sentence and saw two pictures of nouns. They had to match each Czech noun to its picture. Learning of case marking was tested in a grammaticality judgement task. A *d’* analysis revealed that participants were not sensitive to the difference between grammatical and ungrammatical case endings. In Experiment 2, the syntactic patterns of exposure and test sentences were cut down from four to two and an elicited repetition task was added to the exposure phase. However, case marking was not learnt. In Experiment 3, the case markers were reduced to nominative and accusative, the amount of Czech nouns in the exposure sentences was reduced by half, the number of times each case marker appeared in the exposure phase increased and participants were asked to repeat the Czech noun after each exposure trial. Following all these modifications, participants were finally able to distinguish grammatical case endings from ungrammatical ones.

In this chapter’s Experiment 1, not only did the experimental design fail to establish an initial learning effect for the dissimilar structure, in line with Rebuschat (2009) and Rogers, Révész et al. (2015). It also failed to conclusively elicit knowledge of the similar structure, which was shared with the L1. The facilitative role of cross-linguistic syntactic similarity in the initial acquisition of a cross-linguistically similar L2 structure, as opposed to a cross-linguistically dissimilar L2 structure, would be more effectively assessed by an experimental design that clearly elicits and/or triggers learning of the two structures. I addressed this in Experiment 2.

3. Experiment 2

3.1. Overview

Like in Experiment 1, in Experiment 2 I exposed Spanish natives without knowledge of Galician to sentences formed by the similar and the dissimilar structure to then assess the establishment of the two structures in learners' linguistic system. Experiment 2 was identical to Experiment 1 except for four modifications. First, I changed the exposure conditions from incidental to intentional; learners were exposed to the target structures by means of a structure-search task including only plausible sentences. As discussed, the plausibility judgement task in Experiment 1 could have hampered that learners noticed the similar and the dissimilar structure, since plausibility judgements could be made by only paying attention to the verb-noun combination at the end of each sentence. By directing learners' attention towards the form of the sentences, I aimed to facilitate noticing of the structures and, hence, processing and acquisition (cf. Rebuschat, 2009). Second, I increased the number of sentences in the exposure phase from 100 (50 per structure) to 200 (100 per structure). As mentioned in Rebuschat (2009) and Rogers, Révész et al. (2015), when experiments fail to trigger or detect learning, increasing the amount of input might be helpful. Third, sentences were presented in auditory and written form during exposure (and during the test). In Experiment 1, presenting sentences aurally could have caused that learners had difficulty segmenting the infinitive verb in the dissimilar structure separate from the following article, instead processing the verb as a finite form. By presenting sentences in auditory and written form, I expected segmentation and, thus, acquisition, to be facilitated. Finally, learners were tested on their learning of the structures in a GJT *with feedback*. Previous studies have shown that feedback might facilitate L2 syntax learning (e.g. Carroll & Swain1993; Doughty & Varela, 1998; Leeman, 2003; Long et al., 1998; Mackey & Philp, 1998; Muranoi, 2000; Rosa, 1999). Hence, I expected that, if the exposure phase did not elicit knowledge of the similar structure or trigger learning of the dissimilar one, feedback would. The experiment ended with a verbal report.

The hypotheses for Experiment 2 were identical to the ones for Experiment 1. That is, I hypothesized that the similar structure would be processed using an L1 structure present in the linguistic system by the time that L2 acquisition started, but that the dissimilar structure would have to be learnt during the experiment. If, in spite of all the modifications, Experiment 2 did not result in learning of the dissimilar structure, then **Hypothesis 1 (H1)** claimed that the similar structure would be established in learners' linguistic system, but the dissimilar structure would be not. Alternatively, **Hypothesis 2 (H2)** claimed that the similar and the dissimilar structure would both be established in learners' linguistic system, but the former would be more robustly established than the latter.

The changes to Experiment 1's design additionally posed some secondary, methodological questions. Given the introduction of feedback in the GJT, **Methodological question 1 (MQ1)** asked, *"If the dissimilar structure becomes established in learners' linguistic system, will this occur during the exposure phase or during the testing phase, as a result of feedback?"*. **MQ1's Hypothesis 1 (MQ1_H1)** claimed that the dissimilar structure would become part of the system during the exposure phase; **MQ1's Hypothesis 2 (MQ1_H2)** claimed that the dissimilar structure would become part of the system during the testing phase. On the other hand, if the similar and the dissimilar structure were both part of learners' grammar by the end of the exposure phase, **Methodological question 2 (MQ2)** asked, *"Will feedback increase learning of the similar and the dissimilar structure during the test?"*. In MOGUL terms, I propose that, since feedback indicates whether a structure is appropriate in the L2 or not, it could cause that the resting activation level of the similar and the dissimilar structure increased and, hence, that they became more firmly established in learners' linguistic system (Sharwood Smith & Truscott, 2014, sec. 4.3). However, for this to be possible, the resting level of the two structures at the beginning of the test should be at a point that allows raising. MOGUL assumes that, if the resting activation level of a representation is extremely high, this will not be further raised by additional processing (Sharwood Smith & Truscott, 2014, sec. 4.6.5). Since learners would process the dissimilar structure for the first time in the exposure phase, I hypothesized that its resting activation level would not be extremely high and, thus, it could be raised in the GJT. By contrast, learners processed the similar structure prior to Experiment 2, as part of the L1. Depending on learners' experience with the structure, its resting activation level could be extremely high, in which case it would not be further raised by feedback. Considering this, I postulated two hypotheses. **MQ2's Hypothesis 1 (MQ2_H1)** claimed that learning of the similar and the dissimilar structure would increase because of feedback. Consequently, if by the end of the exposure phase the similar structure was more robustly established in the linguistic system than the dissimilar one, this would not vary throughout the test. **MQ2's Hypothesis 2 (MQ2_H2)** claimed that learning of just the dissimilar structure would increase because of feedback. Consequently, the similar structure could be more robustly established in the linguistic system than the dissimilar one by the end of the exposure phase but, by the end of the testing phase, both structures could be comparably established in learners' grammar.

3.2. Participants

Forty-four Spanish natives (35 female) participated in the study. Their ages ranged from 18 to 35 (M = 21, SD = 3.44). They were all studying at the University of the Basque Country (UPV/EHU). All reported no previous knowledge of Galician or other languages with inflected infinitives in the linguistic background questionnaire filled out prior to the experiment (the same as in Experiment 1, see Appendix A-1). More than half of participants (59%, 26/44) reported having some knowledge of Basque (proficiency at or under B2 level). Yet, they all declared feeling more comfortable using Spanish. Additionally, the 44 participants reported using only Spanish with their parents prior to starting school (0-3 years). The questionnaire asked about the frequency with which participants used Spanish and Basque on a 7-point scale (1 = *Spanish only*, 7 = *Basque only*) during childhood (3-12 years), puberty (12-18 years) and adulthood (after 18 years) at school/university/work, at home and at other places. The mean scores for each environment in each period of life are available in Appendix A-2. Spanish was the most used language throughout participants' lives (mean language use during childhood, 1.45 (*SD* = 1.01); puberty, 1.44 (*SD* = 1.01); adulthood, 1.21 (*SD* = 0.49)). Additionally, participants self-rated their proficiency speaking, listening, reading and writing in Spanish on a 7-point scale (1 = *very poor proficiency*, 7 = *perfect proficiency*) and considered this was at a near perfect level (mean proficiency level across skills, 6.76 (*SD* = 0.53)). The mean scores for each skill can also be consulted in Appendix A-2. All participants reported having normal or corrected to normal vision and hearing. Before the experiment began, they read and signed an informed consent (Appendix A-3). Experiment 2 was part of the project "Cross-linguistic activation effects in bilingual language processing and learning" (PGC2018-097970-B-I00), funded by the Spanish Ministry of Science, Innovation and Universities and approved by the Committee of Ethics for research involving human beings of the University of the Basque Country (*Comité de Ética para las Investigaciones con Seres Humanos, CEISH,* Ref. M10_2019_167). Participants received 7€ for their participation.

3.3. Materials

The mini-language used in Experiment 1, consisting of a Spanish-Galician cognate vocabulary and adapted Galician syntax, was also used in Experiment 2. The object of study were the similar and the dissimilar structure detailed in Section 2.3.

3.3.1. Exposure set

Since the exposure phase consisted of a structure-search task involving only plausible sentences, the exposure set in Experiment 2 only included the plausible constructions in Experiment 1's exposure set (see Section 2.3.1). These were 100 sentences formed by the similar structure and the corresponding 100 sentences formed by the dissimilar structure. To prevent participants from hearing and reading the two versions of a given sentence, two lists were created, each containing 50 instances of the similar structure and 50 of the dissimilar structure. The sentence recordings used in Experiment 1 were also used in Experiment 2. The lexicon of the exposure set included the same five impersonal expressions as in Experiment 1, the same 20 verbs, the same 20 proper nouns and the same 20 inanimate singular nouns used in plausible sentences (the 20 inanimate nouns used in implausible sentences were eliminated, see Appendix A-4 for the complete list of vocabulary). A given impersonal expression occurred with a specific proper noun, verb and inanimate noun only once in the exposure set. This can be consulted in Appendix A-9.

3.3.2. Testing set

The testing set was identical to that in Experiment 1 (Section 2.3.2) and, thus, included four conditions: the grammatical similar structure, the grammatical dissimilar structure, the ungrammatical similar structure and the ungrammatical dissimilar structure. There were 80 sentences per condition. The four versions of each sentence were divided into four lists, containing 80 items each (20 instances of the grammatical similar structure, 20 of the ungrammatical similar structure, 20 of the grammatical dissimilar structure and 20 of the ungrammatical dissimilar structure). The sentence recordings in Experiment 1 were used in Experiment 2. The lexicon of testing set was identical to the one in Experiment 1 (five impersonal expressions, 20 verbs, 20 proper nouns and 20 inanimate singular nouns). The verbs and nouns of the exposure and testing sets were different (see Appendix A-4). The combination of a given impersonal expression, proper noun, verb and inanimate noun occurred only once in the testing set. This is available in Appendix A-9.

3.4. Procedure

Participants were told that they would participate in a study on sentence comprehension in Galician. The experiment consisted of three phases: an exposure phase (Section 3.4.1), a testing phase (Section 3.4.2) and a debriefing phase (Section 3.4.3). It was conducted using the E-Prime 2.0 software. Participants were tested individually in a soundproof booth. Sentences were played through headphones. The experiment was conducted in Spanish and it had a maximum duration of 45 minutes. The instructions are reported in Appendix A-10.

3.4.1. Exposure phase

Participants were informed that they would read and listen to some sentences in Galician, half having a structure and the other half having a different structure. They were instructed to pay attention to the form of the sentences and to try to find out the two structures according to which they could be formed. Additionally, they were told that, in a second part of the experiment, they would be tested on their knowledge of the two structures, i.e. they would read and listen to new sentences and would have to decide whether these were formed by the same structures as the sentences in the first part of the experiment or not. Each sentence in the exposure set was presented twice, so each participant read and listened to 200 sentences (100 formed by the similar structure and 100 formed by the dissimilar structure). Sentences were played one by one and in a randomized order for each participant. As each sentence played, it appeared written at the centre of a white computer screen. Immediately after the audio of a sentence ended, the following sentence was automatically presented. The exposure phase lasted for around 10 minutes.

3.4.2. Testing phase

Participants were reminded that they had read and listened to sentences formed by two structures, and that they would read and listen to novel sentences in Galician and would have to indicate, as quickly as possible, whether these were formed by one of the structures of the language or not. Mirroring Experiment 1, learners were taught that those sentences that were formed by the structures of the language were to be judged as “correct” and those that were not, as “incorrect”. Sentences were presented in a randomized order for each participant.

Participants listened to the sentences on an item-by-item basis and each sentence also appeared written at the centre of a white computer screen. Once the audio of a sentence ended, this was replaced by the options *Correcto* ("Correct") and *Incorrecto* ("Incorrect") and their associated response keys. The key "A" had to be pressed to indicate that the sentence was correct and the key "L" to indicate that it was incorrect. The two response options appeared on the screen for 5 seconds. If after this time participants had not chosen one of the two, a message informed that no response was detected and a new sentence was automatically presented. If participants pressed "A" or "L", feedback on the accuracy of the response was shown for 700ms (a green tick if participants were right, a red cross if they were wrong). Then, a new sentence was played (Figure 2.4). The testing phase was preceded by a short practice session to familiarize participants with the task. This included four sentences that were not part of the testing set (a grammatical similar structure, a grammatical dissimilar structure, an ungrammatical similar structure and an ungrammatical dissimilar structure).

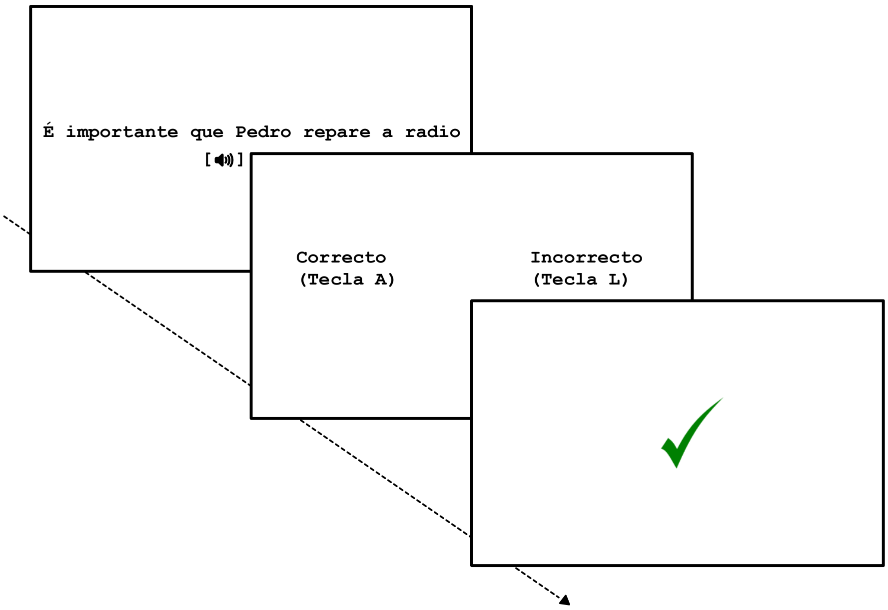


FIGURE 2.4. Example test trial in Experiment 2. Participants first read and listened to a sentence. The speaker between brackets (not shown in the experiment) indicates that the sentence was aurally presented. When the audio of the sentence ended, the words *Correcto* ("Correct") and *Incorrecto* ("Incorrect") appeared on the screen, together with the response key for each option, *Tecla A* ("Key A") or *Tecla L* ("Key L"). After pressing one of the keys, feedback was provided for 700ms.

3.4.3. Debriefing phase

At the end of the experiment, participants were encouraged to verbalize any knowledge about the similar and the dissimilar structure. They were asked the same questions as in Experiment 1, repeated below for convenience:

1. As mentioned during the experiment, the sentences you listened to were formed according to two structures. Did you notice which structures were these?
2. If yes, please indicate which you think these structures are.

The questions appeared on the screen accompanied by a text box in which participants typed their answers. Since learning would take place under intentional conditions, it was expected to be explicit and to result in conscious, verbalizable knowledge of the similar and the dissimilar structure (cf. Chapter 1, Section 1.3). Thus, participants were expected to provide metalinguistic descriptions of the structures (see this chapter, Section 2.4.3 for examples of descriptions).

3.5. Predictions

In Section 3.1, I described my hypotheses for the main research question and the secondary, methodological questions of the experiment. The **main research question** asked whether cross-linguistic syntactic similarity would facilitate initial L2 syntax acquisition, as evidenced in greater learning of a cross-linguistically similar L2 structure compared to a cross-linguistically dissimilar one. I hypothesized that, while the former would be processed using the equivalent L1 structure, which would be a very stable item in the linguistic system, the latter would have to be learnt during the experiment. If, despite the modifications from Experiment 1 to Experiment 2, learning of the dissimilar structure did not occur, then I hypothesized that the similar structure would be established in learners' linguistic system, but the dissimilar structure would be not **(Hypothesis 1, H1)**. Alternatively, if the dissimilar structure was learnt, I hypothesized that the similar and the dissimilar structure would both be established in learners' linguistic system, but that the former would be more firmly established than the latter **(Hypothesis 2, H2).** Like in Experiment 1, whether one of these hypotheses was correct or not will be indicated by learners' performance in the GJT, assessed by accuracy and *d'* analyses. My predictions for H1 and H2 were the same as in Experiment 1 (see Section 2.5).

Methodological question 1 (MQ1) asked whether learning of the dissimilar structure, if any, would occur during the exposure phase or during the testing phase, as a result of feedback. Naturally, I proposed two hypotheses: the dissimilar structure would become part of the linguistic system during the exposure phase **(MQ1_H1)** or during the testing phase **(MQ1_H2).** While my experimental design did not allow fully disentangling between these two hypotheses, the closest way in which this could be done was by examining learners' performance immediately after the exposure phase, i.e. in the first trials of the GJT. I focused on participants' first 20 test trials[18]. Like in Experiment 1, I considered that a structure was learnt if learners accepted grammatical sentences formed by that structure and rejected

[18] The minimum number by which all learners had classified instances of the two grammatical and the two ungrammatical structures. Overall number of trials per condition: grammatical similar structure, 233; ungrammatical similar structure, 201; grammatical dissimilar structure, 217; ungrammatical dissimilar structure, 216.

ungrammatical sentences violating it (accuracy analysis) or if learners were sensitive to the difference between that structure and its ungrammatical counterpart (*d'* analysis). **If MQ1_H1 was correct** and the dissimilar structure was part of learners' linguistic system by the end of the exposure phase, I predicted that in the first trials of the GJT learners would already accept the dissimilar structure and reject its ungrammatical counterpart significantly above chance. Additionally or alternatively, they would be sensitive to the difference between the two structures. By contrast, **if MQ1_H2 was correct** and the dissimilar structure became part of the linguistic system during the testing phase, I predicted that overall performance in the GJT would reveal learning of the dissimilar structure, but learners' performance in the first trials of the test would not.

Finally, **Methodological question 2 (MQ2)** asked whether feedback would increase learning of the structures during the GJT. A first hypothesis **(MQ2_H1)** maintained that learning of the similar and the dissimilar structure would increase because of feedback. Consequently, if by the end of the exposure phase the similar structure was more robustly established in the linguistic system than the dissimilar one, this would not vary throughout the test**.** Alternatively, a second hypothesis **(MQ2_H2)** maintained that learning of just the dissimilar structure would increase because of feedback. Consequently, the similar structure could be more robustly established in the linguistic system than the dissimilar one by the end of the exposure phase but, by the end of the testing phase, both structures could be comparably established**.** To test these hypotheses, I compared learners' performance in the first and the last 20 trials of the GJT. **If MQ2_H1 was correct,** I predicted that accuracy for the grammatical and the ungrammatical similar structure, and the grammatical and the ungrammatical dissimilar structure, or sensitivity to the difference between the two pairs of constructions, would be higher in the last test block than in the first one. Additionally, in the two blocks accuracy for the grammatical and the ungrammatical similar structure, or sensitivity to the difference between the two, would be higher than for the grammatical and the ungrammatical dissimilar structure. **If MQ2_H2 was correct,** I predicted that accuracy for the grammatical and the ungrammatical dissimilar structure, or sensitivity to the difference between the two, would be higher in the last test block than in the first one. Yet, accuracy for the grammatical and the ungrammatical similar structure, or sensitivity to the difference between the two, would be comparable in the two test blocks. Additionally, accuracy for the grammatical and the ungrammatical similar structure, or sensitivity to the difference between the two, would be higher than for the grammatical and the ungrammatical dissimilar structure in the first test block, but not in the last one.

3.6. Coding and data analysis

I used the programming environment R (R Core Team, 2022, version 4.2.2) to analyse the data of the experiment. The function and package used to compute each statistical test and effect size measure are only cited if not mentioned in Experiment 1.

3.6.1. Testing phase

The accuracy of the response in each trial of the GJT was coded as binary (1 = correct grammaticality judgement, 0 = incorrect grammaticality judgement). Trials in which participants could not make a judgement in the time allowed (5 seconds) were removed. Overall, 0.38% (17/3520) of all trials were excluded: 0.57% (5/880) of grammatical similar structures, 0.57% (5/880) of grammatical dissimilar structures, 0.23% (2/880) of ungrammatical similar structures and 0.57% (5/880) of ungrammatical dissimilar structures.

The similar and the dissimilar structure as part of the linguistic system

I started by assessing whether the test provided evidence that the similar and the dissimilar structure were part of learners' linguistic system. This was done in the same way as in Experiment 1. First, I descriptively examined mean accuracy percentages for the grammatical similar structure, the ungrammatical similar structure, the grammatical dissimilar structure and the ungrammatical dissimilar structure. Percentages were compared against chance (50%) via one-sample t-tests. Next, to determine whether participants had a tendency to accept or reject sentences in the GJT, I coded responses as Hits, False alarms, Misses or Correct rejections. Then, I calculated the mean index of response bias *c* for the grammatical and the ungrammatical similar structure, on the one hand, and the grammatical and the ungrammatical dissimilar structure, on the other hand. I compared these indices against zero using one-sample t-tests. A significant response bias was found, so I calculated mean *d'* scores reflecting sensitivity to the difference between the two pairs of grammatical-ungrammatical structures separated from response bias. These were compared against chance (zero) using one-sample t-tests. As will be shown in Section 3.7.1, the accuracy and *d'* analyses indicated that the similar and the dissimilar structure were established in learners' linguistic system. Hence, to assess whether the dissimilar structure was learnt during exposure or during the test, I examined learners' performance in the first 20 trials of the GJT. I conducted the same analyses that I did to examine whether overall the two structures were part of learners' grammar.

Comparing the establishment of the similar and the dissimilar structure in the linguistic system

To determine whether, as hypothesized, the similar structure was more firmly established in learners' linguistic system than the dissimilar structure, I analysed the effect of cross-linguistic similarity on accuracy. I compared accuracy for the grammatical and the ungrammatical similar structure vs. the grammatical and the ungrammatical dissimilar structure. Thus, I collapsed accuracy for a grammatical structure and its ungrammatical counterpart, in line with previous studies using GJTs to measure learning of L2 syntactic rules or patterns (e.g. Kim & Fenn, 2020; Morgan-Short et al., 2014; Rogers, Révész, et al., 2015; Tagarelli et al., 2016). Since accuracy is a binary variable, I used a generalized linear mixed effects model. The model had Accuracy as a dependent variable, Cross-linguistic similarity as a fixed effect, random intercepts by participant and item and a by-participant random slope of Cross-linguistic similarity, to account for the fact that the effect of this variable could differ for each participant. Treatment coding was used for the variable Cross-linguistic similarity, so that the category *Similar structure* was assigned the value 0 and the category *Dissimilar structure* was assigned the value 1. Following the accuracy analysis, I calculated two *d'* scores for each participant, one reflecting sensitivity to the difference between the grammatical and the ungrammatical similar structure and the other reflecting sensitivity to the difference between the grammatical and the ungrammatical dissimilar structure. Mean *d'* scores were compared using a paired-samples t-test, since the assumptions of this test were met. Specifically, a Shapiro-Wilk test calculated with the function *shapiro.test* from the *stats* package revealed that the differences between participants' pairs of *d'* scores were not normally distributed ($W = 0.91$, $p = .003$). Yet, since the sample size was larger than 30, this was not a problem for the parametric test (Levshina, 2015). A Levene's test calculated with the function *leveneTest* from the *car* package (Fox & Weisberg, 2019) revealed that the variances in the populations that represented the pairs of *d'* scores were equal ($F\ (1, 86) = 0.23$, $p = .63$)[19].

Next, to assess whether feedback affected learning of the similar and the dissimilar structure from the first 20 trials to the last 20 trials of the GJT, a generalized linear mixed effects model tested for the interaction between the effect of Cross-linguistic similarity (Similar structure vs. Dissimilar structure) and Test block (First vs. Last) on accuracy. Deviation coding was used for the two fixed effects, so that the categories *Similar structure* and *First* were assigned the value 0.5 and the categories *Dissimilar structure* and *Last* were assigned the value -0.5. The model that provided a better fit for the data included random intercepts by participant and by item and by-participant random slopes of Cross-linguistic similarity and Test block, as tested by nested model comparisons[20]. Finally, a 2x2 within subjects repeated-measures ANOVA looked into the interaction between the effect of Cross-linguistic similarity and Test block on *d'* scores. This was calculated using the function *anova_test* from the *rstatix* package. Partial eta-squared (η_p^2) was calculated as a standardized effect size measure for the results yielded by the ANOVA. As suggested by Cohen (1969), I considered a η_p^2 of 0.01, 0.06 and 0.14 to be small, medium and large, respectively.

3.6.2. Debriefing phase

Responses in the verbal report were transcribed and two experimenters classified participants as *aware* or *unaware* of the target structures using the same rubric as in Experiment 1 (see Appendix A-8 for the transcriptions and the rubric). I coded awareness as a binary variable (1

[19] To double check that the parametric test was not affected by the non-normality of the differences between the pairs of *d'* scores, I also performed the non-parametric version of the paired-samples t-test, the Wilcoxon signed-rank test, fitted with the function *wilcox.test* from the *stats* package. The standardised measure of effect size *r* was calculated using *wilcox_effsize* from the *rstatix* package. Following the interpretation in published literature, I considered an *r* of 0.1-0.3, 0.3-0.5 and ≥ 0.5 to be small, medium and large, respectively. The results of the test will be reported in a footnote in the results section.

[20] Comparison of the models with and without a by-participant random slope of the interaction between the two fixed effects: $X^2(4) = 4.72$, $p = .32$.

= aware participant, 0 = unaware participant). Then, I calculated the percentage of aware and unaware participants and I conducted a descriptive analysis of their verbal reports.

3.7. Results

3.7.1. Testing phase

The similar and the dissimilar structure as part of the linguistic system

Table 2.7 shows mean accuracy percentages for the similar structure (SS), the dissimilar structure (DS) and their ungrammatical counterparts (*SS and *DS). Accuracy was above 75% for all conditions. In spite of the high accuracy percentages, the index *c* indicating response bias when judging SS and *SS items, on the one hand, and DS and *DS items, on the other hand, revealed a tendency towards accepting structures rather than rejecting them, irrespective of their grammaticality. In both cases, the mean index was negative and significantly below zero (SS and *SS items, M = -0.16, SD = 0.35; t (43) = -3.01, p < .01, d = -0.15; DS and *DS items, M = -0.10, SD = 0.32; t (43) = -2.01, p = .03, d = -0.30). Thus, I calculated *d'* scores as a measure of sensitivity to the difference between the grammatical and the ungrammatical similar and dissimilar structures unaffected by response bias (Table 2.8). As mentioned in Section 2.6.2, a perfect discrimination between grammatical and ungrammatical stimuli (100% sensitivity) corresponds to a *d'* score of 4.65. The mean *d'* scores reflecting sensitivity to the difference between SS and *SS items and between DS and *DS items were above 2.00, which indicates quite good discriminability. In a nutshell, the accuracy and *d'* analyses show that learners were able to identify that the sentences formed by the similar or the dissimilar structure were part of the L2, but that the sentences violating these structures were not. This is taken as evidence that the similar and the dissimilar structure were part of learners' linguistic system.

	SS	***SS**	**DS**	***DS**
M	89.26***	78.25***	82.29***	76.57***
SD	30.98	41.28	38.20	42.38
95%CI	[87.20, 91.31]	[75.51, 80.98]	[79.75, 84.82]	[73.76, 79.38]

TABLE 2.7. Mean accuracy (%), standard deviations (%) and 95% confidence intervals for all conditions in the GJT of Experiment 2. *M* = Mean, *SD* = Standard Deviation, *95%CI* = 95% Confidence Interval. Significance from chance: *** p < .001.

	SS vs. *SS	DS vs. *DS
M	2.42***	2.06***
SD	1.52	1.58
95%CI	[2.35, 2.49]	[1.99, 2.14]

TABLE 2.8. Mean *d′* scores, standard deviations and 95% confidence intervals reflecting sensitivity to SS vs.*SS and DS vs.*DS in the GJT of Experiment 2. *M* = Mean, *SD* = Standard Deviation, *95%CI* = 95% Confidence Interval. Significance from zero: ***p <.001.

Table 2.9 summarises mean accuracy percentages in the four conditions and mean *d′* scores for SS vs. *SS items and DS vs. *DS items in the first 20 trials of the GJT. As shown, at the beginning of the test accuracy was already significantly above 50% for all conditions and sensitivity to the difference between the two grammatical structures and their ungrammatical counterparts was significantly greater than zero. In short, participants were able to discriminate between grammatical and ungrammatical stimuli immediately after exposure, which indicates that the similar and the dissimilar structure were already established in learners' linguistic system. This suggests that the dissimilar structure was learnt during the exposure phase, in line with Hypothesis 1 postulated for MQ1 (MQ1_H1).

	Accuracy			***d′* scores**		
	M	*SD (%)*	*95%CI*	*M*	*SD*	*95%CI*
SS	85.84***	34.94	[81.33, 90.35]	1.94***	1.14	[1.83, 2.05]
**SS*	73.63***	44.17	[67.49, 79.78]			
DS	78.34***	41.29	[72.82, 83.87]	1.68***	1.39	[1.55, 1.81]
**DS*	69.91***	45.97	[63.74, 76.07]			

TABLE 2.9. First 20 trials of the GJT in Experiment 2: Mean accuracy (%), *d′* scores, standard deviations and 95% confidence intervals for grammatical vs. ungrammatical similar and dissimilar structures. *M* = Mean, *SD* = Standard Deviation, *95%CI* = 95% Confidence Interval. Significance from chance: ***p < .001.

Comparing the establishment of the similar and the dissimilar structure in the linguistic system

Figure 2.5 illustrates mean accuracy percentages when judging the grammatical and the ungrammatical similar structure compared to the grammatical and the ungrammatical dissimilar structure. Learners correctly judged an average of 83.74% (*SD* = 36.91%, 95%CI = [82.01, 85.47]) of all SS and *SS items and 79.43% (*SD* = 40.43%, 95%CI = [77.53, 81.32]) of all DS and *DS items. Accuracy was significantly higher when judging the first pair of structures than the second (β = -0.61, *SE* = 0.20, z = -3.08, p = .002). The estimated coefficient of the effect of Cross-linguistic similarity (Similar structure vs. Dissimilar structure) on accuracy was -0.61 (in log odds). This corresponds to an odds ratio of 0.54 to 1. Thus, the odds of judging a sentence correctly as opposed to incorrectly were 0.54 times smaller when the item was DS or *DS compared to when it was SS or *SS[21]. The results of the accuracy analysis were matched by the *d'* analysis. Figure 2.6 displays the distribution of the *d'* scores reflecting sensitivity to SS vs. *SS items and DS vs. *DS items. Sensitivity to the difference between the grammatical and the ungrammatical similar structure was significantly higher than to the difference between the grammatical and the ungrammatical dissimilar structure, as indicated by a paired-samples t-test (t (43) = 3.06, p = .004, medium effect size of d = 0.5)[22]. In sum, learners identified that the grammatical similar structure was part of the language and that the ungrammatical similar structure was not significantly more accurately than they did so for the grammatical and the ungrammatical dissimilar structures. Likewise, they were significantly more sensitive to the difference between the first pair of structures than the second. This suggests that the similar structure was more robustly established in learners' linguistic system than the dissimilar one, in line with Hypothesis 2 postulated for this chapter's main research question.

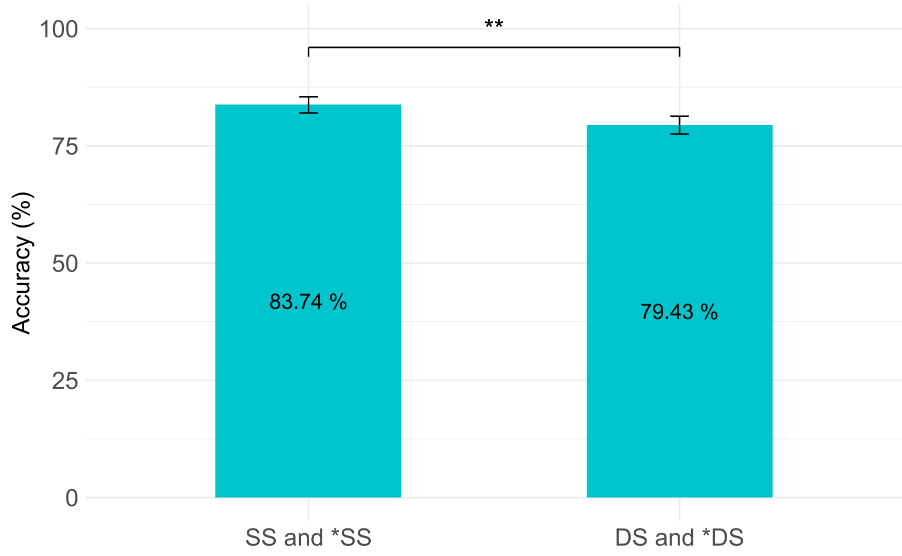


FIGURE 2.5. Mean accuracy (%) for SS and *SS vs. DS and *DS in the GJT of Experiment 2. Error bars represent 95% confidence intervals.

[21] Odds ratio was calculated using the function *exp()* from the R *base* package (R Core Team, 2022).
[22] Non-parametric Wilcoxon signed-rank test: V = 615, p = .002, medium effect size of r = 0.49.

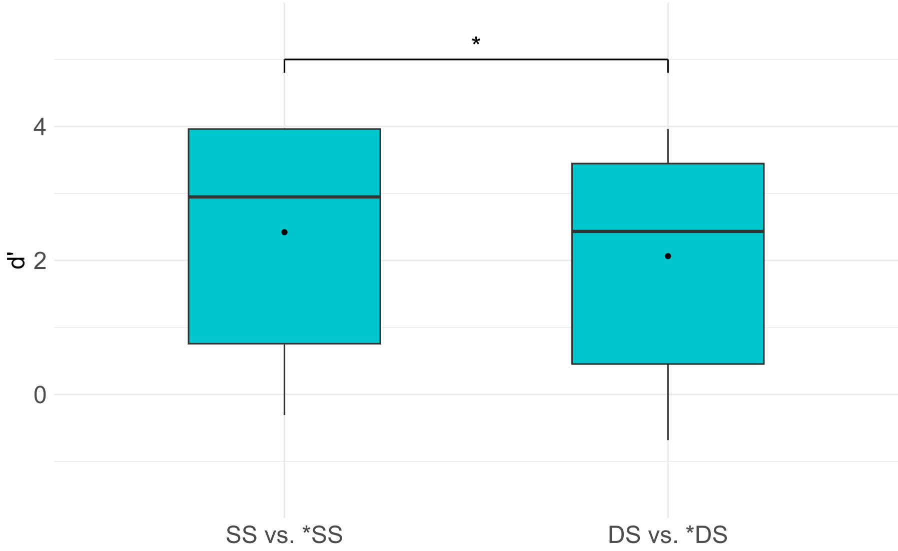


FIGURE 2.6. Distribution of *d'* scores reflecting sensitivity to SS vs.*SS and DS vs.*DS in the GJT of Experiment 2. The black horizontal line in the box of the boxplot shows the median. The black dot is the mean. Whiskers end at the smallest and largest data points that fall within 1.5 times the interquartile range from the first quartile (25%) and the third quartile (75%) of the data.

To determine whether learning of the structures varied throughout the test, I compared learners' ability to discriminate between the grammatical and the ungrammatical similar and dissimilar structures in the first and the last 20 trials of the GJT. Table 2.10 shows mean accuracy percentages and *d'* scores for SS -*SS items and DS -*DS items in the first and the last test blocks. A generalized linear mixed effects model looked into a possible interaction between the effect of Cross-linguistic similarity (Similar structure vs. Dissimilar structure) and Test block (First vs. Last) on accuracy. The model yielded a main effect of Cross-linguistic similarity (β = 0.76, *SE* = 0.17, *z* = 4.39, *p* < .001) and Test block (β = -1.41, *SE* = 0.32, *z* = -4.40, *p* <.001) but no interaction between the two (β = -0.25, *SE* = 0.30, *z* = -0.84, *p* = .40). Additionally, an ANOVA examined whether there was an interaction between the effect of Cross-linguistic similarity and Test block on *d'* scores. Matching the accuracy analysis, the test yielded a main effect of Cross-linguistic similarity (*F* (1, 43) = 9.56, *p* = .003, large effect size of η_p^2 = .182) and a main effect of Test block (*F* (1, 43) = 7.23, *p* = .01, large effect size of η_p^2 = .144), but no interaction between the two variables (*F* (1, 43) = 0.15, *p* = .70, η_p^2 = .004).

		Accuracy			*d'* scores		
		M	*SD (%)*	*95%CI*	*M*	*SD*	*95%CI*
First 20 test trials	*SS - *SS*	80.18	39.91	[76.42, 83.95]	1.57	1.27	[1.45, 1.69]
	*DS - *DS*	74.13	43.84	[69.99, 78.27]	1.22	1.28	[1.10, 1.34]
Last 20 test trials	*SS - *SS*	87.05	33.18	[83.90, 90.20]	1.94	1.14	[1.83, 2.05]
	*DS - *DS*	81.96	38.49	[78.35, 85.58]	1.68	1.39	[1.55, 1.81]

TABLE 2.10. First and last 20 trials of the GJT in Experiment 2: Mean accuracy (%), *d'* scores, standard deviations and 95% confidence intervals for grammatical vs. ungrammatical similar and dissimilar structures. *M* = Mean, *SD* = Standard Deviation, *95%CI* = 95% Confidence Interval.

The main effect of Test block indicates that, overall, accuracy and *d'* scores for the two pairs of structures were significantly higher in the last 20 test trials than in the first 20 test trials. This suggests that the similar and the dissimilar structure became more robustly established in learners' grammar throughout the test. Additionally, the main effect of Cross-linguistic similarity indicates that, in general, accuracy and *d'* scores for SS -*SS items were significantly higher than for DS -*DS items. The lack of a significant interaction between the two variables indicates that the aforementioned effect did not vary between test blocks. This suggests that, by the end of the exposure phase, the similar structure was more firmly established in the linguistic system than the dissimilar one and that this did not vary throughout the test, in line with Hypothesis 1 postulated for MQ2 (MQ2_H1).

3.7.2. Debriefing phase

Participants' verbal reports indicated that 64% (28/44) were *aware* of the similar and the dissimilar structure and the remaining 36% (16/44) were *unaware* of the structures. For all aware participants, the GJT evidenced learning of one or the two structures. Specifically, they all judged SS and *SS sentences and/or DS and *DS sentences at or above 65% accuracy (clearly above chance). Likewise, their *d'* scores reflecting sensitivity to the difference between SS and *SS and/or between DS and *DS were above zero. All aware participants verbalized a description of the similar and the dissimilar structure, except two participants (P19 and P32) who, despite learning the two structures, reported only the similar one. On the other hand, the test evidenced learning of one or the two structures, as defined by the criterion above, for just four out of the 16 unaware participants (P8, P31, P33 and P38). Within these four participants, three commented either on the presence/absence of the complementizer *que* in the two structures (P8), or on the finite/non-finite nature of the embedded verb (P33 and P38). P31 did not answer to the verbal report. Finally, within the 12 unaware participants whose test did not reveal learning of the structures, nine did not verbalize any knowledge in the verbal report and three (P3, P17 and P42) just commented that some sentences contained the word *que* and others did not[23].

3.8. Discussion

In Experiment 2, like in Experiment 1, I explored whether cross-linguistic syntactic similarity had a positive influence at the earliest stage of adult L2 syntax acquisition. I examined whether Spanish natives with no knowledge of Galician showed greater learning of the *similar structure*, shared in Spanish and Galician, compared to the *dissimilar structure,* existing only in Galician. I made several modifications to the paradigm in Experiment 1 seeking that this second experiment effectively elicited knowledge of the similar structure and triggered learning of the dissimilar one. Specifically, I exposed learners to the structures in an auditory *and* visual *structure-search task* including only plausible sentences and *doubling* the number of sentences formed by the similar and the dissimilar structure with respect to Experiment 1. I tested learning in an auditory *and* visual grammaticality judgement task *with feedback*. The experiment concluded with a verbal report. Like in Experiment 1, based on MOGUL (Sharwood Smith & Truscott, 2014) and other theories and models of second language acquisition (e.g. Carroll, 1999, 2001; MacWhinney, 2005; B. D. Schwartz & Sprouse, 1994, 1996; Westergaard, 2021), I assumed that the similar structure would be processed using a construction acquired as part of the L1 and present at the initial state of L2 acquisition. Conversely, the dissimilar structure would have to be learnt during the experiment. Considering this, if both structures ended up being established in learners' linguistic system, I hypothesized that the former would be more firmly established than the latter or, in other words, that there would be a learning advantage for the similar structure over the dissimilar one.

The results of the testing phase provided evidence in favour of this hypothesis. Accuracy when judging the similar structure and its ungrammatical counterpart was significantly higher than when judging the dissimilar structure and its ungrammatical counterpart. Additionally, sensitivity to the difference between the first pair of structures was significantly higher than to the difference between the second pair of structures. This goes in line with previous findings that cross-linguistically similar structures and morphosyntactic features are processed and produced more accurately than cross-linguistically dissimilar ones by complete beginner L2 learners (Tolentino & Tokowicz, 2014) and low-proficient L2 learners (e.g. Bardovi-Harlig, 1997; Chang & Zheng, 2015; C. Ellis et al., 2013; Izquierdo & Collins, 2008; Tokowicz & MacWhinney, 2005; Vainikka & Young-Scholten, 1994). Since several changes were made from Experiment 1 to Experiment 2, it is not possible to pinpoint which modification(s) caused the results obtained, and establishing this is beyond the scope of this chapter. In the next sections, I address how the modifications to the exposure and testing

[23] The results of the testing phase obtained for the whole group of participants replicated when only the subset of participants whose test evidenced learning of one or the two target structures (n = 32) were included in the analysis (see Appendix A-11).

phases could have influenced processing and acquisition by processing of the target structures and I discuss the results of the testing phase and the debriefing phase.

3.8.1. Discussion of the exposure phase

In Experiment 2, contrary to Experiment 1, the exposure phase elicited knowledge of the similar structure and resulted in learning of the dissimilar structure. This is suggested by learners' performance in the first 20 trials of the GJT, which revealed that, immediately after the exposure phase, learners were sensitive to the difference between the grammatical and the ungrammatical similar and dissimilar structures and judged all sentence types significantly above chance. The exposure phase in Experiment 2 differed from the one in Experiment 1 in three aspects, which could explain this result.

First, I changed the exposure task from a plausibility judgement task to a structure-search task. As discussed in Section 2.8.1, it could be that in Experiment 1 participants performed the plausibility judgement task by paying attention only to each sentence's embedded verb and following noun phrase, instead of to the entire sentence. This could have prevented the similar and the dissimilar structure from becoming intake, a necessary step for processing and acquisition by processing (Carroll, 1999; Corder, 1967; Sharwood Smith & Truscott, 2014, sec. 9.3.2; VanPatten, 1996, 2004, and more). In the exposure phase in Experiment 2, learners' attention was directed towards the form of sentences, for they were instructed to try to discover the two structures according to which sentences could be formed. This focus on form could have facilitated that learners noticed the structures and, hence, that they became intake. Second, I presented sentences in auditory and written form (in Experiment 1, sentences were presented just aurally). A second possibility discussed in Experiment 1 was that the dissimilar structure did not become intake and, thus, was not learnt, because Spanish natives processed the infinitive verb (ending in a consonant) and the following article (a vowel) as a single word, misanalysing the infinitive as a finite verb. Presenting sentences aurally and visually in Experiment 2 could have helped participants segment the dissimilar structure as required. Finally, I increased exposure to the similar and the dissimilar structure from 100 to 200 example sentences per structure. A third possibility mentioned in Experiment 1 was that the similar and the dissimilar structure became intake, but that the number of times that the dissimilar structure was processed was not high enough for it to become part of learners' linguistic system. By doubling the number of exposure sentences, the amount of times each structure was processed increased, which could have caused that the dissimilar structure became robustly established in learners' grammar. Irrespective of the explanation(s) assumed, I propose that by the end of the exposure phase, the similar and the dissimilar structure were firmly established in learners' linguistic system, as suggested by performance in the first trials of the test.

In addition, the comparison of learners' performance in the first vs. the last 20 trials of the GJT indicated that accuracy and *d'* scores for the grammatical and the ungrammatical similar structure were significantly higher than for the grammatical and the ungrammatical dissimilar structure, and this did not vary from the first to the last trials of the test. This result suggests that, already by the end of the exposure phase, the similar structure was more firmly established in learners' linguistic system than the dissimilar structure. This finding is consistent with how I hypothesized that complete beginner L2 learners would process and/or acquire cross-linguistically similar and dissimilar structures within the MOGUL framework (Section 1.3). In short, I hypothesize that, when learners were exposed to sentences formed by the similar structure, the equivalent L1 structure activated and was used to process the input. This structure probably had a high resting activation level, since it had been repeatedly processed in the L1. Each time that the similar structure was processed during exposure, its resting activation level raised even more. By contrast, when learners encountered the dissimilar structure, the syntax processor had to create a new representation for it, activating, selecting and combining items in the syntactic store. This representation received a low resting activation level. Each time that the dissimilar structure was subsequently processed, its resting activation level increased, up to the point that it became a stable item in the linguistic system. Crucially, learners' overall experience processing the similar structure was more extensive than processing the dissimilar structure. Consequently, by the end of the exposure phase the first structure had a higher resting activation level and, thus, was more robustly established in learners' linguistic system, than the second. My hypotheses are generally compatible with the Full transfer/Full access model (B. D. Schwartz & Sprouse, 1994, 1996), the Autonomous Induction Theory (Carroll, 1999, 2001), the Unified Competition Model (MacWhinney, 2005) and the Micro-cue model of L2 acquisition (Westergaard, 2021). These approaches all assume that L1 syntax is present at the initial state of L2 acquisition and will be used to parse L2 input. Cross-linguistically dissimilar syntactic aspects will be acquired when the L2 cannot be parsed according to L1 syntax, either by accessing UG or by using a learning or a processing mechanism, a process that is naturally more costly.

3.8.2. Discussion of the testing phase

The main result of the testing phase was that, overall, the similar structure was more firmly established in learners' linguistic system than the dissimilar one. This is indicated by the fact that learners judged the similar structure and its ungrammatical counterpart significantly more accurately than the dissimilar structure and its ungrammatical counterpart, and they were significantly more sensitive to the difference between the former than the latter. I interpret this result as the consequence of (i) exposure to the similar and the dissimilar structure, as described in the previous section, and (ii) any additional learning of the structures that occurred in the testing phase because of feedback. In what follows, I hypothesize how the difference between the two structures in learners' linguistic system would have affected the way that grammatical and ungrammatical structures were processed and judged in the GJT.

I hypothesize that, since the grammatical similar structure was firmly established in learners' minds, it was processed without difficulty and participants readily identified it as part of the L2. This would explain why participants accepted the structure, i.e. considered it "correct" in the L2, on almost 90% of the occasions on average. Turning to the ungrammatical similar structure, I hypothesize that learners processed it as if it was the (grammatical) similar structure until arriving at the embedded verb, where the syntactic violation occurred. Since the similar structure was a very stable item in learners' linguistic system, this could have made it quite easy to identify the mismatch between this structure, containing a finite embedded verb, and the ungrammatical similar structure, containing a non-finite embedded verb. This would explain why participants rejected the ungrammatical structure, i.e. considered it "incorrect" in the L2, on almost 80% of the occasions on average, and why sensitivity to the difference between the grammatical and the ungrammatical similar structure was quite large (mean *d'* score above 2.00). A similar reasoning could be applied to the processing of the grammatical and ungrammatical dissimilar structures, with the difference that, as mentioned, the (grammatical) dissimilar structure was less robustly established than the (grammatical) similar structure in learners' minds. I argue that this could have made it more challenging to identify the dissimilar structure as part of the L2 as well as to identify the mismatch between this structure and its ungrammatical counterpart, which were identical except for the finite/non-finite embedded verb. In sum, this would explain why, overall, accuracy for the grammatical and the ungrammatical similar structure and sensitivity to the difference between the two was significantly higher than for the grammatical and the ungrammatical dissimilar structure.

In addition, the introduction of feedback in the GJT of Experiment 2 allowed assessing whether it increased learning of the target structures by comparing learners' performance in the first vs. the last 20 test trials. Results suggested that this was the case, since accuracy and *d'* scores for the grammatical and the ungrammatical similar structure, on the one hand, and the grammatical and the ungrammatical dissimilar structure, on the other hand, significantly increased from the beginning to the end of the test. The increase was comparable for the two pairs of constructions. This result goes in line with findings from classroom and laboratory studies showing that L2 syntax learners benefit from exposure to feedback (e.g. Carroll & Swain, 1993; Doughty & Varela, 1998; Leeman, 2003; Long et al., 1998; Mackey & Philp, 1998; Muranoi, 2000; Rosa, 1999). Additionally, accuracy and *d'* scores for the similar structure and its ungrammatical counterpart were significantly higher than for the dissimilar structure and its ungrammatical counterpart, and this did not significantly vary from the first test block to the last one. As mentioned, this is consistent with the hypothesis that, by the end of exposure, the similar structure was a more stable item in learners' grammar than the dissimilar structure. Since the two became more robustly established in learners' minds as a result of feedback, by the end of the test the difference between the two remained. I interpret the finding that feedback increased learning of the similar and the dissimilar structure within the MOGUL framework. Yet, other theoretical approaches could also accommodate the positive effect of feedback on L2 development (see Leeman, 2007 for an overview). In short, in Experiment 2 feedback provided information about whether a structure was part of the L2 or not. In MOGUL, if a particular representation is appropriate for language processing, its resting activation level increases. If a representation proves not to be appropriate, its resting activation level does not increase and it does not become part of the linguistic system (Sharwood Smith & Truscott, 2014, sec. 4.3). I hypothesize that since feedback provided information about the appropriateness of a representation for processing, it helped raise the resting activation level of the similar and the dissimilar structure (and not raise the resting activation level of the structures violating them). This increased accuracy and sensitivity to the difference between the two pairs of grammatical-ungrammatical structures[24].

Before concluding, it is worth mentioning that despite the paradigm in Experiment 2 being more successful than that in Experiment 1, the experiment did not elicit knowledge of the similar structure or trigger learning of the dissimilar one for 12 participants. This is not entirely unexpected since, as mentioned in Chapter 1, not all learners are equally successful in learning an L2, and the differences in the level of competence achieved can be explained by diverse factors, including language aptitude, working memory, motivation, and learners' anxiety during L2 learning (see R. Ellis, 2004; S. Li et al., 2022). Determining which of these factors was responsible for the results obtained is beyond the scope of this chapter.

3.8.3. Discussion of the debriefing phase

The verbal report in the debriefing phase indicated that 28 out of the 32 participants whose GJT evidenced learning of the similar and/or the dissimilar structure could verbalize a metalinguistic description of them. Thus, they were considered *aware* of the structures. All aware participants verbalized the similar and the dissimilar structure, with the exception of two participants who reported only the similar structure. The GJT evidenced that these two participants had knowledge of both structures. In fact, they said that they did not report the dissimilar structure due to an error when conducting the verbal report. They pressed the key ENTER to describe the dissimilar structure in a separate line in the text box where participants wrote their answers but, in doing this, the verbal report automatically ended. The remaining four participants who learnt the structures were considered *unaware* of them; one of these participants did not verbalize any knowledge and, the other three, provided incomplete descriptions of the structures. It is not clear whether these participants did not verbalize a (more accurate) description of the similar and the dissimilar structure because they did not have explicit knowledge of them or because they lacked the ability to put their knowledge into words. In addition, as mentioned in Experiment 1 (Section 2.8.3), it could be that the learner who did not verbalize a response had explicit knowledge of the structures but was not confident enough or simply chose not to answer to the verbal report (Rebuschat, 2013). From these results, it is clear that most participants who learnt the structures developed verbalizable knowledge of them. Participants were exposed to the structures under

[24] In MOGUL, if a representation has an extremely high resting activation level, additional processing will not raise it any more (Sharwood Smith & Truscott, 2014, sec. 4.6.5). This could have been the case for the similar structure, which learners had processed before as part of the L1. Yet, since accuracy and sensitivity to the difference between the grammatical and the ungrammatical similar structure increased throughout the test, it seems that the resting activation level of the similar structure at the beginning of the testing phase, in spite of being high, still allowed some raising.

intentional learning conditions, by means of a structure-search task and then a grammaticality judgement task with feedback. Hence, this result goes in line with previous studies showing that L2 syntax learning under intentional conditions results in conscious knowledge (e.g. N. C. Ellis, 1993; Rebuschat, 2009; Robinson, 1997; Tagarelli et al., 2016). Finally, the GJT did not evidence learning of the structures for 12 participants, who were *unaware* of these structures according to their verbal reports. On the one hand, nine did not answer to the verbal report. This can be simply attributed to the fact that, since no learning of the structures occurred, participants could not verbalize them. On the other hand, the remaining three participants reported that some sentences included *que* and others did not, but did not mention that in some sentences the embedded verb was conjugated and, in others, it was not. Like in Experiment 1, this can be attributed to the fact that *que* was probably more salient than the subjunctive and infinitive verb endings and, hence, easier to report (see Section 2.8.3).

4. Concluding remarks

This chapter sheds new light on one of the factors facilitating initial L2 syntax acquisition, in particular the facilitative role of cross-linguistic syntactic similarity in the acquisition of L2 syntactic structures by adult complete beginner learners, which is an under-researched topic. It contributes to our understanding of the availability of L1 syntax for L2 processing from the start of the acquisition process, as well as of how real-time processing may lead to the acquisition of a syntactic structure that is not similar in the L1 and the L2 from learners' first exposure to the language. Based on the MOGUL framework (Sharwood Smith & Truscott, 2014) and other approaches to L2 acquisition (e.g. Carroll, 1999, 2001; MacWhinney, 2005; B. D. Schwartz & Sprouse, 1994, 1996; Westergaard, 2021), I hypothesized that learners would initially attempt to process L2 input using L1 syntax. When this was not possible, the linguistic system would have to be adapted to accommodate the L2. Thus, I predicted that after an initial exposure to a cross-linguistically similar and a cross-linguistically dissimilar L2 structure, learners would show a learning advantage for the similar structure, which would be part of the grammar from the start, over the dissimilar structure, which would have to be learnt from input. In Experiment 1, various shortcomings of the experimental design, arguably including the characteristics of the exposure and testing phases, prevented learning of the dissimilar structure and did not effectively elicit knowledge of the similar structure. Following the modification of the experimental paradigm in Experiment 2, my prediction was met. In sum, this chapter shows that cross-linguistic syntactic similarity might be beneficial at the earliest stage of L2 syntax acquisition, but also that appropriate exposure and testing conditions are crucial for capturing such facilitation.

Chapter 3

The facilitative role of lexical frequency in initial L2 syntax acquisition

1. Introduction

Word frequency affects lexical processing efficiency, so that high frequency words are processed faster and more accurately than low frequency words both in the L1 and in the L2. This is known as the *word frequency effect* (WFE). Word frequency is closely related to features such as word length, age of acquisition and similarity to other lexical items (Brysbaert et al., 2018). Yet, megastudies such as the one conducted by Brysbaert et al. (2016) have provided evidence for an effect of word frequency on lexical processing isolated from other potentially confounding variables. In Chapter 2, I showed that cross-linguistic syntactic similarity between the L1 and the L2 facilitates initial L2 syntax acquisition. In Chapter 3, I investigate whether differences in lexical frequency facilitate initial L2 syntax acquisition, focusing on cross-linguistically similar and dissimilar structures.

Several studies have found that high frequency words are recognized, named and read faster and more accurately than low frequency words, as evidenced in a wide range of lexical processing tasks in the L1 and in the L2. These include, amongst others, lexical decision tasks (e.g. L1, Dupoux & Mehler, 1990; Monsell et al., 1989; L1 and L2, Duyck et al., 2008; Gollan et al., 2011), (picture-) naming tasks (e.g. L1, Balota & Chumbley, 1984; Forster & Chambers, 1973; L1 and L2, Gollan et al., 2008; Van Wijnendaele & Brysbaert, 2002) and reading tasks using eye-tracking (e.g. L1, Inhoff & Rayner, 1986; Rayner & Raney, 1996; L1 and L2, Cop et al., 2015; Whitford & Titone, 2012). In addition, ERPs have shown that high frequency words elicit lower amplitudes than low frequency ones in time windows associated with lexical activation, which also suggests that the former are accessed more easily than the latter (e.g. Dufour et al., 2013; Hauk & Pulvermüller, 2004; Rugg, 1990). To determine whether a word is to be considered high frequency or low frequency, researchers usually look at its frequency of occurrence in the language, as measured by objective frequency counts. These are obtained by consulting corpora based on different types of texts, such as books, newspapers or magazines (e.g. Baayen et al., 1995; Kučera & Francis, 1967), television subtitles (e.g. Brysbaert & New, 2009; Cuetos et al., 2011), social media (e.g. Gimenes & New, 2016) or a combination of written and oral texts (e.g. Davies, 2015; Real Academia Española, 2021). The most complete corpora are those containing frequency measures from both written and oral documents, for they take into account that written frequency may differ from spoken or heard frequency (Balota et al., 2001). To cope with the fact that frequency counts may vary with corpus size, researchers use a standardized frequency measure, e.g. frequency per million words or frequency measured in the Zipf scale (Brysbaert et al., 2018). Frequency of occurrence can also be assessed by means of subjective familiarity ratings, i.e. by asking participants to grade words on a scale according to their familiarity. However, objective frequency counts tend to be favoured over subjective familiarity ratings, since the latter may be influenced by other factors besides familiarity, such as the age of participants, the number of meanings of a word or the knowledge of its orthographic to phonological correspondences (Balota et al., 2001).

Objective frequency counts reflect the number of times that a form has been used or has been encountered in the language. However, that a word has been encountered in the input is no guarantee that it has become intake and has been processed. From a theoretical point of view, it is possible to distinguish between frequency of exposure and the frequency with which a form appears in the mental representation of the input constructed during processing. Within the MOGUL framework, these two types of frequency are referred to as *external frequency* and *internal frequency,* respectively (Sharwood Smith & Truscott, 2014, sec. 4.6.5). This distinction is also relevant within other processing approaches acknowledging the difference between input and intake (e.g. Carroll, 1999). Whereas external word frequency is a measure derived from a sample of selected texts and is shared for all the speakers of a language, internal word frequency may vary from one individual to another, depending on the particular input to which each person has been exposed and whether or not the words in this input have become intake. As will be detailed in the next section, the measure ultimately influencing lexical processing is internal frequency, i.e. frequency of occurrence as coded in the mental representation of words. Internal frequency is a characteristic of words in the mind; it is not a number that can be directly accessed and used as an experimental variable in research. Nevertheless, as Speelman and Krisner (2005) suggest, there is a direct relation between internal frequency and external frequency. For instance, if a word classified as high frequency based on its frequency of occurrence in a corpus is recognized faster than a word classified as low frequency, then it could be inferred that the former not only has a higher external frequency than the latter, but also a higher internal frequency. In Chapter 3, I will assume external frequency to be equivalent to internal frequency (henceforth simply referred to as *lexical frequency*).

To my knowledge, no study has investigated whether and how differences in lexical frequency, as measured by objective frequency counts, affect L2 syntax learning. At most, some have examined whether differences in lexical frequency affect syntactic processing in the L1 and the L2 and have yielded mixed results (Hopp, 2016; Luoni, 2022; Staub, 2011; Tily et al., 2010, see Section 1.2 for the details). To investigate this, Experiment 3 replicated Experiment 2, in which Spanish natives with no knowledge of Galician learnt Spanish-Galician similar and dissimilar embedded clauses. Yet, Experiment 3 manipulated the lexical frequency of the cognate embedded verbs in the L1, so that it was lower than in Experiment 2. As far as I know, no model, theory or framework of L2 acquisition makes explicit predictions about how processing cross-linguistically similar and/or dissimilar structures with high frequency verbs, as opposed to low frequency verbs, may affect the initial acquisition of these structures. I propose that hypotheses about this may be derived within the MOGUL framework, considering how it suggests that high frequency and low frequency words are stored and processed in the linguistic system and how lexical and syntactic processing interact during syntax acquisition (see Section 1.3 for a full account). In short, I propose that high frequency verbs should facilitate the acquisition of the cross-linguistically dissimilar structure, which has to be learnt from input (i.e. there should be a learning advantage for participants in Experiment 2 over participants in Experiment 3). Conversely, the facilitation should be smaller, or even non-significant, for the cross-linguistically similar structure, which is assumed to be established in the linguistic system at the beginning of L2 acquisition (i.e. there should be a smaller or non-significant learning advantage for participants in Experiment 2 over participants in Experiment 3). In addition, the learning advantage for the cross-linguistically similar structure over the dissimilar one found in Experiment 2 with high frequency verbs, should replicate in Experiment 3, with low frequency verbs. Yet, the effect should be larger for structures processed with low frequency verbs than with high frequency verbs. The comparison of the results of Experiment 2 and Experiment 3 provided evidence in favour of these hypotheses.

Chapter 3 is organized as follows. In Section 1.1, I review some studies evidencing the word frequency effect and I describe how this could be explained in some of the most influential models of monolingual and bilingual word processing, including MOGUL. Next, in Section 1.2, I discuss the influence that lexical frequency may have on syntax processing, a necessary step for understanding the influence that frequency may have on syntax acquisition. In Section 1.3, I detail how I propose that lexical frequency might affect the initial acquisition of cross-linguistically similar and dissimilar L2 structures within MOGUL. Then, in Section 2, I report and discuss Experiment 3, comparing its results to those of Experiment 2. The chapter ends with some concluding remarks in Section 3.

1.1. The influence of lexical frequency on word processing

The advantage of high frequency words over low frequency words is prevalent in visual and spoken word recognition, comprehension and naming in the L1 and the L2, both in isolation and in context (e.g. in a sentence or a paragraph). In the monolingual domain, there is a wide range of models of visual and spoken word recognition and/or comprehension, examples being the logogen model (Morton, 1969), the Interactive Activation model (McClelland, 1987; McClelland & Elman, 1986; McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982), the Cohort model (Gaskell & Marslen-Wilson, 1997; Marslen-Wilson, 1987, 1990), and more. Likewise, several models of word production have been proposed (e.g. Dell, 1986, 1990; Levelt et al., 1999). As mentioned in Chapter 1, in the bilingual domain the most influential model of word recognition and comprehension, with possible extensions to production, is the Bilingual Interactive Activation plus model (BIA+, Dijkstra & van Heuven, 2002), which is an adaptation of the Interactive Activation model to bilingualism. All these models have integrated the word frequency effect in their architecture. I propose that the MOGUL framework can also explain the WFE in monolingual and bilingual word processing, and quite similarly to the BIA+ model. In what follows, I review some L1 and L2 studies exemplifying the WFE. Since Chapter 3 is concerned with syntax acquisition occurring as a result of visual and aural sentence processing, I will focus on the WFE in visual and spoken word recognition and processing, both in and out of context. Given the similarity of MOGUL with the (bilingual) Interactive Activation model, I will detail how the WFE can be accounted for in the Interactive Activation model and the BIA+ model. Finally, I will discuss the similarities and differences between the BIA+ model and MOGUL and how the latter could explain the WFE.

One of the first studies looking into the role of lexical frequency in word processing was Howes and Solomon (1951). In two experiments, English natives were exposed to high frequency and low frequency words tachistoscopically, i.e. starting with a very short presentation (30ms) which progressively lengthened until recognition. Participants had to report the word they thought they saw after each presentation. Overall, briefer exposure durations were required to report correctly high frequency words compared to low frequency words. Further evidence for the WFE emerged in the decade of the 70s, the 80s and the 90s, as experimental methodology developed. For instance, Monsell et al. (1989) showed that lexical frequency affected visual lexical decision time and semantic and syntactic categorization time. In a first experiment, English natives classified high, medium and low frequency nouns as denoting a person or a thing. In addition, participants performed a visual lexical decision task using the target nouns and a series of non-words. Word frequency significantly affected lexical decision times and classification times for person and thing nouns, so that they decreased as frequency increased. In a second experiment, another group of English natives performed a visual lexical decision task using the same materials as in Experiment 1. Additionally, they conducted a syntactic (noun/adjective) categorization task using the same nouns as in the first task and a set of high, medium or low frequency adjectives. The results of Experiment 2 matched those of Experiment 1, i.e. the higher word frequency was, the faster participants responded to the lexical decision task and the faster they categorized person nouns, thing nouns and adjectives. Another relevant study is that by Dupoux and Mehler (1990), who investigated the WFE in an auditory lexical decision task. French natives listened to high frequency and low frequency words and non-words and were asked to distinguish the former from the latter. High frequency words were identified significantly faster than low frequency words and elicited significantly fewer errors. Additional evidence for the WFE was found in other studies using the visual lexical decision task (e.g. Besner & McCann, 1987; Blosfeld & Bradley, 1981; Hudson & Bergman, 1985; Norris, 1984; Paap et al., 1987; Scarborough et al., 1977; Whaley, 1978), the auditory lexical decision task (e.g. Blosfeld & Bradley, 1981; Connine et al., 1990; Marslen-Wilson, 1987) and the semantic categorization task (e.g. Forster & Shen, 1996).

As this research was being conducted, some studies tested whether the WFE held when words were not presented in isolation, but processed as part of a sentence. An example is the study by Inhoff and Rayner (1986), who had English natives read NP-V-NP sentences differing in the frequency of the noun in the first NP (e.g. *the slow music/waltz captured her attention*, p. 432). Participants read the sentences for comprehension while their eye-movements were recorded. There were three conditions: a) a full line condition, during which participants saw the whole sentence, b) a two-word window condition, during which participants only saw the word being fixated, the words to the left and a word to the right and c) a one-word window condition, during which participants only saw the word being fixated and the part of the sentence to its left. Overall, first fixation durations (the time spent on a word the first time it is looked at) were shorter on high frequency nouns than on low frequency nouns. In addition, low frequency nouns were more often refixated than high frequency ones, causing gaze durations (the sum of all fixations before the eyes move to another word) to be significantly longer for the former than for the latter. The parafoveal preview of high frequency words in the full line condition and the two-word window condition caused first fixations on these words to be shorter than on low frequency words. In short, this study showed that high frequency words were processed faster than low frequency words also when embedded in a sentential context. Similar results were reported by Rayner and Duffy (1986); Just and Carpenter (1980); Rayner and Raney (1996); Rayner et al. (1996) and Schilling et al. (1998).

As advanced, one of the models accounting for the WFE in the L1 is the Interactive Activation model, which is the basis for the BIA+ model. The Interactive Activation model was initially proposed to account for visual word recognition (McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982), but a model of spoken word recognition based on the architecture of the Interactive Activation model was developed and tested a few years later (the TRACE model of speech perception, McClelland & Elman, 1986). These models propose that the processing system is divided into several levels: the visual feature level, the letter level and the orthographic word form level, for visual perception, and the acoustic feature level, the phoneme level and the phonological word form level, for speech perception. Each level contains a series of units (features, letters/phonemes or words) linked to units in the adjacent levels. For simplicity's sake, in what follows I discuss McClelland and Rumelhart's model of visual word recognition, but a similar reasoning should hold for spoken word recognition.

In the absence of input, features, letters and words are in an inactive state, referred to as the *resting activation level* and conceptualized as activation being at or below zero. When a written word is encountered, features compatible with it activate and activation spreads to letters containing those features via the connections between the feature level and the letter level. At the same time, those letters that do not contain the features in the input are inhibited. The same process occurs from the letter level to the word level, i.e. active letters spread their activation to words containing them and inhibit words that do not contain them. All active letters or words compete to be interpreted as the letter or the word in the input, inhibiting one another. The processing system is more likely to report that the input includes the letter or the word that is most active. Processing occurs in cascade: when a unit in a given level is activated, activation immediately spreads to the adjacent level. That is, the processing system does not wait for activity in a level to be complete before spreading activation to the next. Instead, all levels process the input in parallel. Additionally, perception is conceived as an interactive process (hence, the name *Interactive* Activation model). This means that the flow of activation not only operates in a bottom-up manner, from the feature level up to the word level, but also in a top-down manner: activation of a word in the word level (for instance, due to contextual information) spreads down to the letter and the feature levels, influencing the perception of its letters and features. The two processes occur simultaneously.

To account for the word frequency effect, this model assumes that the resting activation level of a word is shaped by how frequently it has been activated in the past. Simply put, when a word is encountered, the activation of compatible units in each level jumps from the resting level to a positive value (above zero), causing those units to become active. This triggers the recognition process described. Once a word has been recognized, its activation falls back towards the resting level, landing slightly above the original one. This causes that high frequency words have higher resting activation levels than low frequency words, i.e. that the resting activation level of the former is at zero, or closer to zero, than that of the latter. Crucially, since the resting level of high frequency words is higher than that of low frequency words, the former become active faster and, hence, are recognized faster. Some years after the Interactive Activation model was proposed, McClelland (1987) developed the Interactive Activation framework, including a syntactic and a word-sense level of representation. In this way, the framework could account for how words were recognized *and comprehended*. That is, how visual or spoken input was incrementally matched onto the orthographic or phonological forms of candidate words and how the best fit was selected, but taking into account that the syntactic and semantic information of a word had to be accessed to process it and to integrate it into a sentence. Crucially, if high frequency words are accessed faster than low frequency words, assuming that each orthographic and phonological word form is connected to a syntactic and a semantic representation means that the syntactic and semantic representations of high frequency words would also be accessed faster than those of low frequency words. This would explain the WFE in the studies on sentence reading and on syntactic and semantic categorization reviewed.

A final important notion concerns the relation between frequency and resting activation level. As mentioned, within the Interactive Activation model the resting activation level of words is at zero or below zero, depending on their frequency of occurrence. The resting activation level increases each time that a word is used and the higher the resting level is, the faster a word is accessed. A natural consequence of this reasoning is that, if two words are encountered very often, at some point their resting activation level will rise no more (i.e., it will be at zero for the two). Hence, even if one of the words is more frequent than the other one, both might be accessed equally fast (B. Gordon & Caramazza, 1985; see also Griffin & Bock, 1998; Kirsner & Speelman, 1996).

Some years after the formulation of the Interactive Activation model, the Bilingual Interactive Activation (BIA) model of visual word recognition was proposed (Dijkstra et al., 1999; Grainger & Dijkstra, 1992). As in the Interactive Activation model, in the BIA model the processing system comprised a feature level, a letter level and an orthographic word form level of representation. Later, the model was extended into the BIA+ model (Dijkstra & van Heuven, 2002), which incorporated phonological and semantic levels of representation. Syntactic representations were not explicitly included but, according to the authors, they could be assumed (cf. Dijkstra & van Heuven, 2002, p. 186). The BIA+ model assumes that visual word processing occurs as described for the Interactive Activation model. However, since lexical access is language non-selective, when a bilingual sees a word, active letters at the letter level activate compatible orthographic word forms in the two languages of the bilingual. These orthographic word forms, in turn, spread their activation to phonological, (syntactic) and semantic representations of words in the two languages. All representations active at a given level compete and inhibit each other regardless of the language to which they belong. Each orthographic and phonological word form is connected to a *language node*, so that when a word is active, the corresponding language node is active as well. At the same time, a given language node inhibits competing lexical representations from the other language to modulate cross-linguistic interference. Importantly, as advanced in Chapter 1, in the BIA+ model frequency is coded in the same way as in the Interactive Activation model, i.e. in the resting activation level of representations. Accordingly, the bilingual model explains the WFE in the same way as its monolingual counterpart, as high frequency words having higher resting activation levels than low frequency words, consequently becoming active faster and being processed faster. The BIA+ model was conceived as a model of bilingual visual word recognition and comprehension, but the authors suggested that it could be generalized to spoken word recognition and comprehension, given that auditory lexical access is also language non-selective (Marian & Spivey, 1999; Schulpen et al., 2003).

In the 2000s, some studies showed that the WFE was present in L2 word recognition and comprehension. To name a couple of examples, Duyck et al. (2008) investigated how Dutch-English bilinguals performed a visual lexical decision task with high frequency and low frequency words in Dutch and in English. Lexical frequency differed within languages, but was matched across languages. Results showed that decision time was significantly faster for high frequency words than for low frequency words both in the L1 and in the L2. Similarly, classification accuracy was significantly higher for high frequency words than for low frequency words in the two languages. Gollan et al. (2011) had English natives, highly proficient Spanish-English bilinguals and less proficient Dutch-English bilinguals conduct a visual lexical decision task and a sentence reading task in English (bilinguals' L2). In the first task, participants had to distinguish between high or low frequency nouns and non-words. The English natives and the two groups of bilinguals performed the task faster and more accurately for high frequency words than for low frequency words. In the second task, the high frequency and low frequency nouns were embedded in two types of sentences, one in which the high/low frequency word was highly predictable and one in which it was not.

Participants read sentences for comprehension while their eye-movements were recorded. Monolinguals and bilinguals obtained shorter gaze durations on high frequency words than on low frequency words irrespective of the type of sentence in which they occurred.

In sum, these studies show that the WFE in visual word processing also occurs in the L2, in isolation and in context. Similar results can be found in visual lexical decision tasks (e.g. Brysbaert et al., 2017; de Groot et al., 2002), progressive demasking tasks (e.g. Lemhöfer et al., 2008) and in reading tasks using eye-tracking (e.g. Cop et al., 2015; Whitford & Titone, 2012)[1]. At the beginning of this section, I claimed that the MOGUL framework could also account for the WFE in the L1 and the L2, in a way that highly resembles that in the BIA+ model. This was advanced in Chapter 1 and is further described in the next section.

The influence of lexical frequency on word processing within MOGUL

MOGUL and the BIA+ model share many representational and functional characteristics. MOGUL also assumes that lexical access is language non-selective and that there is an integrated lexicon for words of the two languages of the bilingual. However, as Sharwood Smith and Truscott claim: "there is no lexicon in MOGUL in the traditional sense" (2014, p. 250). Instead of conceiving the lexicon as a separate entity in the linguistic system, MOGUL proposes that each module has its own lexicon. Lexical items (from the L1, the L2 or other) are conceived as chains of coindexed representations stored in the different sublexicons (Sharwood Smith & Truscott, 2014, sec. 2.3.3). As such, for each word there is an acoustic representation in the auditory module, an orthographic representation in the visual module, a phonological representation in the phonological module, a syntactic representation in the syntactic module and a semantic representation in the conceptual module. Each representation is connected to representations in the immediately preceding and following modules. Yet, instead of direct connections between representations, as in the BIA+ model, MOGUL proposes interfaces connecting coindexed representations in adjacent modules (Sharwood Smith & Truscott, 2014, sec. 1.7.1).

Like the BIA+ model, MOGUL claims that all representations have a resting activation level. However, it does not postulate specific numbers for this level. When a speaker reads or listens to a word, orthographic and/or phonological representations compatible with it from all the languages a speaker knows activate. More specifically, activation raises from the resting activation level up to the current activation level, which is the sum of the resting level and the activation received during the on-going processing (Sharwood Smith & Truscott, 2014, sec. 3.3.5). Activation spreads from the orthographic and/or phonological representations to coindexed syntactic and semantic representations in adjacent modules. All active

[1] Perhaps because the claims of the BIA+ model have only been fully developed for bilingual visual word processing, experimental studies have focused on visual word processing in and out of context, overlooking the WFE in bilingual auditory word processing (Murao & Kajiro, 2017). Yet, if the architecture of the BIA+ model can be extended to spoken word recognition, as Dijkstra and van Heuven (2002) mention, similar frequency effects could be expected also in this field.

representations compete for selection by the processor in each module and the most active representation is selected. As in the BIA+ model, in MOGUL processing operates in cascade, i.e. as soon as there is activation in a module this spreads to adjacent modules (Sharwood Smith & Truscott, 2014, sec. 3.3.4). In addition, activation also flows bottom-up and top-down simultaneously. For instance, when the semantic representation of a word is activated, it sends activation back to its syntactic, phonological and orthographic representations in the corresponding modules (Sharwood Smith & Truscott, 2014, sec. 3.4.4).

Unlike in the BIA+ model, in MOGUL language nodes are unnecessary. Sharwood Smith and Truscott propose the conceptual triggering hypothesis, according to which, instead of language nodes, there is a conceptual representation in the conceptual module associated with each of the languages of a bilingual. This conceptual representation is directly linked to acoustic and orthographic word representations via an interface between the auditory/visual modules and the conceptual module. Hence, each time that the acoustic or orthographic representation of a word activates, the conceptual representation of the target language activates as well. At the same time, the acoustic or visual representation of a word activates the corresponding phonological, syntactic and conceptual representations via the corresponding interfaces. The phonological and syntactic modules process the input without being “aware” of the language being processed (Sharwood Smith & Truscott, 2014, sec. 6.4.3).

Importantly, in MOGUL, just as in the BIA+ model, frequency is coded in a representation’s resting activation level. When a chain of representations is stimulated (for instance, when reading or listening to a word), its activation jumps from the resting level to the current activation level. When stimulation fades, activation progressively declines towards the resting level, stopping somewhere above it. If a word has just been processed for the first time, the chain of representations constituting it will have a very low resting activation level. Each time that the chain is active, its resting activation level will increase (Sharwood Smith & Truscott, 2014, sec. 3.3.5). Considering all this, I assume that high frequency words will be stored in the linguistic system with a higher resting activation level than low frequency words. Additionally, as mentioned, the current activation level of a representation is equivalent to its resting activation level plus the activation received during the processing event. I propose that, if the resting activation level of high frequency words is higher than that of low frequency words, when activation from the current processing is added, the former will have a higher current activation level than the latter, i.e. high frequency words will be more strongly activated than low frequency words (see Figure 3.1).

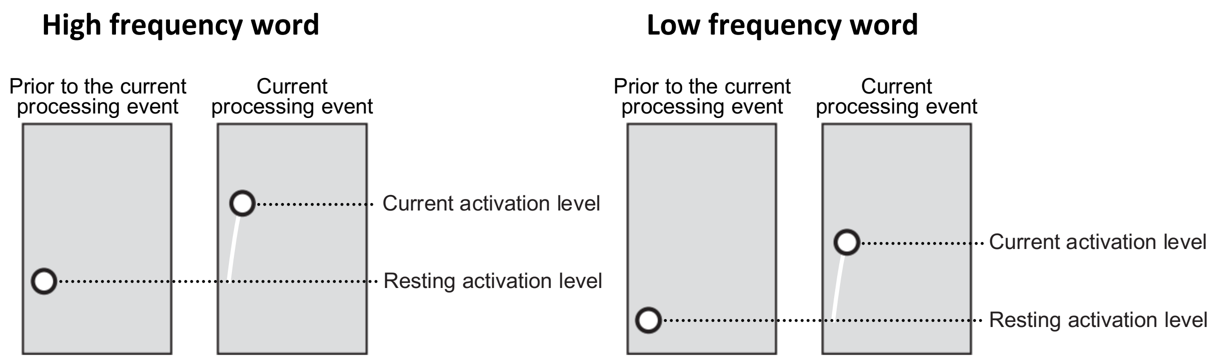


FIGURE 3.1. Example of resting activation level prior to the current processing event and current activation level during the processing event for high and low frequency words.

The fact that the chains of representations constituting high frequency and low frequency words have different resting and current activation levels would directly explain the WFE in the L1 and the L2. As advanced, MOGUL assumes that during processing multiple representations activate in parallel and compete for selection by the processor in each module. Whether a representation is ultimately selected depends on its current activation level: the most active representation amongst all competitors is the one selected. If the chain of representations constituting a high frequency word has a higher current activation level than the chain of representations constituting a low frequency word, then it is more likely that the former will be the most active candidate among competitors and, thus, that it will be selected for processing. In addition, representations with a high resting activation level become active faster than representations with a low resting activation level (Sharwood Smith & Truscott, 2014, sec. 3.4.1). Hence, high frequency words will also be available for selection faster than low frequency words. All this would explain why high frequency words are recognized and processed faster and more accurately than low frequency words.

1.2. The influence of lexical frequency on syntax processing

During language comprehension, we often encounter words in sentences. To comprehend the input, we must recognize and process these words and integrate them into a syntactic structure, which has to be processed as well. In the previous section, I discussed lexical processing, understood as the retrieval of the orthography and/or phonology, syntax and semantics of a word, as well as the effect that frequency has on this process. Since Chapter 3 investigates whether and how lexical frequency affects L2 syntax acquisition, it is necessary to understand first whether and how lexical frequency affects syntax processing. In spite of the relevance of this question for understanding language comprehension, lexical processing and syntactic processing have traditionally been studied separately. As Tily et al. (2010) pointed out:

> Lexical retrieval is typically investigated using individual words presented out of context or in simple carrier phrases, while work on syntactic processing difficulty usually treats lexical retrieval only as a source of confound, explicitly controlling for correlates of lexical access difficulty (p. 913).

In line with this, the models of word processing accounting for the WFE reviewed in the previous section do not specify how a (high frequency or low frequency) word is integrated in the syntactic context of the sentence it belongs to, nor how exactly this syntactic context may influence lexical processing. As for models of sentence processing, they mention that lexical retrieval is the starting point of syntactic structure building. However, in describing how lexical information is accessed, they do not take into account that words differ in frequency of occurrence. As a matter of example, consider the most popular models of sentence processing: garden-path models and constrained-based lexicalist models.

Garden-path models, also referred to as syntax-first models, assume that the processing system works in two stages (e.g. Frazier, 1987, 1989; Frazier & Clifton, 1997; Frazier & Fodor, 1978; Frazier & Rayner, 1982). In the first stage, the system incrementally retrieves the grammatical category of lexical items. This displays the phrases permitted by the grammar, which are used to formulate a single syntactic analysis or parse (e.g. a verb can combine with a complement to form a VP, a determiner and a noun can be combined into a DP). If the grammatical categories encountered are compatible with more than one analysis, the processing system resorts to the Minimal Attachment principle, according to which new input must be attached to the syntactic structure using the fewest number of nodes possible. In other words, the simplest syntactic alternative is chosen. If two analyses are comparably simple, the processing system turns to the Late Closure principle, which dictates that, if possible, new material must be attached to the clause or phrase being processed (Frazier & Rayner, 1982). In the second stage, lexical-semantic properties are accessed and integrated into the analysis (but see Friederici, 2002 for a model that separates access to these properties and integration of syntactic and semantic information into two different phases). If the parse adopted in the first stage proves inadequate given the semantic information available in the second stage, a reanalysis takes place and a new syntactic structure is constructed. Garden-path models can account for the finding that, when reading sentences such as *Since Jay always jogs a mile seems like a short distance to him,* speakers first analyse *a mile* as if it were the direct object of *jogs* (i.e. they attach *a mile* to the phrase being processed, Late closure principle). Later, when the rest of the sentence is processed and it no longer makes sense syntactically to attach *a mile* to the verb *jogs* (when a garden-path effect is obtained), semantics are used to reanalyse the parse and to interpret *a mile* as the subject of the clause following the verb (Frazier & Rayner, 1982). Evidence for garden-path effects has been found in the L1 (e.g. Ferreira & Clifton, 1986; Ferreira & Henderson, 1991; Meng & Bader, 2000) and the L2 (e.g. Jackson, 2008; Jacob & Felser, 2016; Juffs, 1998). Crucially, despite explaining some experimental findings, garden-path models do not discuss the role that differences in a word's frequency of occurrence may play in parsing.

Turning to constraint-based lexicalist models (e.g. MacDonald et al., 1994; McClelland et al., 1989; Trueswell & Tanenhaus, 1994), they differ from garden-path models in that they do not hypothesize separate phases in processing. Instead, all forms of information or "constraints" (general syntactic biases, word category, lexical-semantic information, and more) are accessed simultaneously and are immediately used to parse sentences. Constraint-based models assume that processing is interactive; all types of information are used to process the input, influencing one another to arrive at the most appropriate parse. Furthermore, these models believe that, if the input is ambiguous between two or more syntactic analyses, all alternative parses activate in parallel and compete for selection by the processing system. These analyses can be more or less activated depending on the number of constraints they are consistent with as the input is incrementally processed. Ultimately, the parse that is most compatible with the available constraints is selected. Interestingly, these models discuss access to frequency information related to lexical items. For instance, they acknowledge that some verbs take certain complements more frequently than others (e.g. *forget* is followed more frequently by an object NP than by a sentential complement, cf. Trueswell et al., 1993). Likewise, the same verb form may correspond to two different tenses (e.g. *searched* may be a past participle or a past tense) but it may be more frequently used as one of the two (cf. Trueswell, 1996). All this information is assumed to be encoded in lexical items and to guide parsing (for this reason, constraint-based *lexicalist* models).

Evidence in favour of a lexically-driven parse has been found in L1 and L2 studies (e.g. L1, MacDonald et al., 1994; Trueswell, 1996; Trueswell et al., 1993; Trueswell & Tanenhaus, 1994; L2, Dussias & Cramer Scaltz, 2008; Lee et al., 2013). For instance, Dussias and Cramer Scaltz (2008) had English natives and Spanish-English bilinguals read temporarily ambiguous English sentences with verbs having a direct object bias (e.g. *confirm*) or a sentential complement bias (e.g. *admit*) and followed by either a direct object or a sentential complement. When the verb had a direct object bias, it took both groups of participants significantly longer to read the disambiguating region of a sentential complement compared to that of a direct object (e.g. *The CIA director confirmed the rumor could mean a security leak* vs. *The CIA director confirmed the rumor when he testified before congress,* p. 3). By contrast, when the verb had a sentential complement bias, both groups read the disambiguating region of a sentential complement faster (e.g. *The ticket agent admitted the mistake might not have been caught,* p.3). This indicates that the subcategorization preferences of verbs were accessed and used to resolve the temporary ambiguity. Nevertheless, just as syntax-first models, constraint-based lexicalist models do not incorporate lexical frequency as described in Section 1.1, nor discuss whether and how it may interact with syntactic processing.

As far as I know, two studies have shown that there are links between lexical frequency and processing of syntactic structures (Hopp, 2016; Tily et al., 2010). Tily and colleagues had English natives conduct a word-by-word self-paced reading experiment involving subject-extracted and object-extracted cleft sentences with high frequency or low frequency verbs (1). On the one hand, low frequency words take longer to process than high frequency words (cf. Section 1.1). This is a lexical effect. On the other hand, object-extracted relative clauses and cleft sentences are more challenging to comprehend than subject-extracted ones, as reflected by differences in processing time in the embedded verb (e.g. Ford, 1983; Gennari & MacDonald, 2008; P. C. Gordon et al., 2001, 2002; Traxler et al., 2002, 2005). Although this has been attributed to multiple causes, including the greater difficulty of reanalysing a subject-extracted structure as an object-extracted structure (Clifton & Frazier, 1989) or of integrating long-distance dependents (Gibson, 1998), most explanations agree that the difficulty of processing object-extracted relative clauses and clefts resides in a structure building or structure selection process (i.e. it is a structural effect).

(1) a. Subject-extracted cleft with high frequency verb:
It was Vivian who lectured Terrence for always being late

b. Object-extracted cleft with high frequency verb:
It was Vivian who Terrence lectured for always being late.

c. Subject-extracted cleft with low frequency verb:
It was Vivian who chided Terrence for always being late.

d. Object-extracted cleft with low frequency verb:
It was Vivian who Terrence chided for always being late.

(Tily et al., 2010, pp. 913–914)

Tily et al. divided the subject and object-extracted clefts into different regions, including a *cleft region* (the embedded NP and the verb, e.g. “lectured Terrence” or “Terrence lectured” in 1a and 1b) and a *post-cleft region* (the two words after the cleft region). The authors predicted that, if lexical processing influenced syntactic processing, the lexical effect and the structural effect would interact. That is, when sentences contained a high frequency verb, lexical retrieval would be fast and the larger structural cost of processing object-extracted clefts compared to subject-extracted ones would be observed, as usual, around the embedded verb (in the cleft region). By contrast, when sentences contained a low frequency verb, lexical retrieval would be slower and the structural process underlying the difficulty of processing object-extracted clefts would be delayed until the post-cleft region. Conversely, if lexical and syntactic processing were independent, the lexical and the structural effect would not interact; the difference in processing subject vs. object-extracted clefts would be observed in the same region for sentences with a high or a low frequency verb. The results confirmed the first set of predictions. Overall, subject-extracted clefts were read faster than object-extracted clefts and high frequency verbs were read faster than low frequency verbs. The difference in reading times between subject and object-extracted structures emerged in the cleft region with high frequency verbs and in the post-cleft region with low frequency verbs.

Tily et al. claimed that none of the existing models of sentence processing predicted exactly their results. Yet, they argued that the influence of lexical processing on syntactic processing observed resonated with constraint-based lexicalist models, in the sense that, also in these models, lexically encoded information guides syntactic parsing. The authors interpreted their results as evidence that lexical retrieval of the verb must take place before the start of the structure building or selection process underlying object-extracted clefts. As such, when lexical retrieval time increased by manipulating lexical frequency, the structural process was deferred. Nevertheless, Tily and colleagues acknowledged that lexicalist models do not predict that structural processes must start after lexical retrieval is complete. In fact, if processing is interactive and all sources of information are used optimally to process the input, syntactic processing should also influence lexical processing (cf. McClelland, 1987). For example, most constraint-based lexicalist models assume that, as words are incrementally processed, one or more syntactic structures are constructed in parallel and lexical items associated with these structures are anticipated or predicted (McRae & Matsuki, 2013).

A few years later, Hopp (2016) replicated the study by Tily et al. (2010) with a group of English natives and a group of German natives with intermediate to advanced proficiency in L2 English. The structures studied were also subject and object-extracted clefts containing either a high frequency or a low frequency verb. Importantly, Hopp piloted a study using Tily et al.'s materials and found that most low frequency verbs were unknown to English learners. Consequently, verbs were adapted through a norming study. The high frequency and low frequency verbs chosen for the experiment were significantly more frequent than the high frequency and low frequency verbs in Tily et al.'s study. The two groups of participants read the four conditions during a self-paced reading task. Reading times in the cleft and the post-cleft region were analysed. On the one hand, the results of the native group diverged from those of Tily et al.'s native speakers. In Hopp's experiment, the cleft region in subject-extracted clefts was read significantly faster than that in object-extracted clefts *in conditions with low frequency verbs* (in Tily et al. this was the case in conditions with high frequency verbs). By contrast, reading times for the two structures were comparable across regions in conditions with high frequency verbs. The difference between studies was attributed to low frequency verbs in Hopp's experiment being roughly equivalent in frequency to high frequency verbs in Tily et al.'s experiment. This arguably caused Hopp's natives to process sentences with low frequency verbs as Tily et al.'s natives processed sentences with high frequency verbs. On the other hand, the results of the L2 group did mirror those of Tily and colleagues. That is, in conditions with high frequency verbs, the cleft region in subject-extracted structures was read significantly faster than the same region in object-extracted structures. Conversely, in conditions with low frequency verbs, the advantage in reading times for subject-extracted clefts over object-extracted clefts surfaced in the post-cleft region. Hopp interpreted this performance in the same way as Tily et al. did, i.e. as evidence that lexical retrieval of the verb preceded the structure building or selection process in object-extracted clefts. Therefore, when lexical access was slowed down by low frequency verbs, the structural effect normally surfacing at the verb was delayed and emerged later in the sentence. Hopp linked this result to his Lexical Bottleneck Hypothesis (first mentioned in Hopp, 2014; fully developed in Hopp, 2018), which claims that slowdowns in lexical processing can lead to delayed or incomplete syntactic processing. However, as its name indicates, this is a hypothesis. None of the existing models of sentence processing would account for these results.

In sum, the studies by Tily et al. (2010) and Hopp (2016) provide evidence that differences in lexical processing, as modulated by frequency, may influence syntactic processing in the L1 and the L2. However, other studies have failed to find this effect. For instance, aiming to expand Tily et al. and Hopp's findings to a language other than English, Luoni (2022) investigated the effect of lexical frequency on processing Italian subject and object-extracted clefts. Italian natives and L2 Italian learners listened to the two types of structures in sentences with a high frequency or a low frequency verb and were asked to repeat each sentence exactly as it was heard. Italian natives were equally accurate repeating subject and object-extracted clefts and structures with high frequency and low frequency verbs. Italian learners repeated subject-extracted clefts and structures with high frequency verbs significantly more accurately than object-extracted clefts and structures with low frequency verbs. In neither of the two groups did lexical frequency and structure type interact. Similarly, Staub (2011) conducted a number of eye-tracking experiments looking into how English natives processed sentences with temporary object-subject ambiguities disambiguated by a high frequency or a low frequency verb (e.g. *The boss heard the manager remembered/suppressed some inconvenient facts,* p. 431). Low frequency verbs obtained longer first-pass fixations than high frequency verbs. The syntactic ambiguity elicited more regressions to the verb or longer fixations on subsequent words compared to unambiguous control sentences. However, the lexical and the structural effect did not interact.

More research has to be conducted before the relation between lexical processing and syntactic processing can be fully understood. Crucially, if differences in lexical frequency affect the parsing of syntactic structures, this would need to be explained by a sentence-processing model or framework that incorporates the findings of lexical processing research, including the word frequency effect. I propose that an appropriate framework may be the MOGUL. On the one hand, as detailed in Section 1.1, it describes how lexical items are stored and processed in the linguistic system while taking into account that words differ in frequency of occurrence. On the other hand, it describes sentence processing in real time (see Chapter 1, Section 3.2). Importantly, MOGUL also details how lexical processing interacts with syntax processing during language comprehension. In what follows, I describe this with the help of an example and I detail how I propose that differences in lexical frequency could influence syntactic processing within this framework. Since MOGUL assumes that there is a single processing system for the L1 and the L2 (Sharwood Smith & Truscott, 2014, sec. 3.5), the reasoning should be valid for the two languages.

The influence of lexical frequency on syntax processing within MOGUL

I will focus on the processing activity that occurs in the syntactic module due to the relevance of syntactic processing for syntax acquisition, which is the ultimate focus of this chapter (remember that, in MOGUL, acquisition is simply "the lingering effects of processing", Sharwood Smith & Truscott, 2014, p. 93). However, the process described would also occur in parallel in the phonological module and in the conceptual (semantic) module. That syntactic information is processed separately from semantic information resonates with syntax-first models. Nevertheless, the seriality of these models (the fact that a syntactic parse is constructed in a first stage, autonomously from semantics, and that this only influences syntactic processing in a later stage) is incompatible with MOGUL. Within this framework, there are no separate stages in processing; in line with constraint-based lexicalist models, processing is interactive and the representations constructed in the different modules are simultaneously available and influence each other. Semantic information notably constraints the syntactic representation constructed (Sharwood Smith & Truscott, 2014, sec. 3.3.7, 3.4.1).

Taking this into account, consider how a VO construction such as *kick the ball* would be processed within MOGUL (Sharwood Smith & Truscott, 2014, sec. 4.5.3). This was already advanced in Chapter 1 (Section 3.6) and is further detailed here. Upon listening to the word *kick,* its phonological representation would activate in the phonological store. Activation would spread in a bottom-up manner to the syntactic representation [V_i] coindexed with it in the syntactic store. Activation of [V_i] would spread to the verb's subcategorization frame [V_i $NP_{p, q, r\ldots}$] and to other syntactic representations containing [V_i][2]. These representations would activate in parallel and would compete for selection by the syntactic processor as input is incrementally processed. In addition, activation would weakly spread from [V_i $NP_{p, q, r\ldots}$] to the syntactic representations of NPs and nouns linked with the verb *kick*, maybe including *the ball* and *ball.* As this is occurring, the determiner *the* would be processed. Its phonological representation would activate the syntactic representation [D_x], which would spread its activation to phrases including it, such as [$_{NP}$ D_x $N_{b, c, d\ldots}$]. Activation would spread from [$_{NP}$ D_x $N_{b, c, d\ldots}$] to particular representations of nouns. These phrases and nouns may or may not be the same that were weakly activated by the verb's subcategorization frame. Meanwhile, *ball* would be processed. The syntactic representation [N_b] would activate and would be merged with [D_x] to form [$_{NP}$ D_x N_b]. The NP would spread its activation to other structures containing it, further activating [$_{VP}$ V_i NP_b] and causing it to be selected by the syntactic processor amongst any alternative representations active. The NP (*the ball*) would be combined with [V_i] (*kick*) to form a VP (*kick the ball*).

Now imagine that, instead of processing *kick the ball,* the construction to be processed was *puncture the ball. Kick* has a higher frequency of occurrence than *puncture* (59.21 vs. 1.43 lemmatic frequency per million, Corpus of Contemporary American English, Davies, 2015). To understand how differences in lexical frequency would influence syntactic processing in this framework, two aspects have to be considered. First, for a word to be available for processing as part of a syntactic structure, its syntactic representation (e.g. [V], [D] or [N]) has to be

[2] The subindices in all syntactic representations exemplified here and throughout this chapter are arbitrary and identify the chain of representations corresponding to a specific word, such as *kick*. Following Chomsky (1982), Sharwood Smith and Truscott do not include the subject in the verb's subcategorization frame. Additionally, they prefer to use the label NP instead of DP (2014, p. 45).

activated. This may occur via two routes, 1) as activation spreads from the phonological representation of the word to its syntactic representation and 2) as activation spreads from a larger syntactic representation containing it to the very same syntactic representation of the word. In any case, the extent and the speed of the activation will depend on the representation's resting activation level (Sharwood Smith & Truscott, 2014, sec. 4.5.3). As mentioned in Section 1.1, the chain of representations constituting a high frequency word would have a higher resting activation level than the chain of representations constituting a low frequency word. This would cause that the former is activated faster than the latter. Consequently, the syntactic representation of the high frequency verb (*kick*) would be available for processing in a syntactic structure faster than the syntactic representation of the low frequency verb (*puncture*).

Second, the fact that the chain of representations constituting a high frequency word has a higher resting activation level than that constituting a low frequency word causes that its current activation level is also higher. This is because, as discussed, a representation's current activation level is the result of adding its resting activation level and any activation received during the on-going processing (Sharwood Smith & Truscott, 2014, sec. 3.3.5) (see Figure 3.1). The fact that high frequency words have a higher current activation level than low frequency words has important consequences for syntactic processing. As explained, activation spreads from a syntactic representation such as [V], [D] or [N] to other representations containing it, i.e. their current activation level is raised. Crucially, the degree of the rise is influenced by the current activation level of the representation that activation spreads from (Sharwood Smith & Truscott, 2014, sec. 3.4.1). In short, the higher its current activation is, the higher the rise in the current activation of the representation including it. That being so, the syntactic structure containing a high frequency word should experience a higher rise in current activation level than the syntactic structure containing a low frequency word. That is, the representation $[_{VP}\ V_i\ NP_b]$, corresponding to *kick the ball,* should have a higher current activation level than the representation $[_{VP}\ V_j\ NP_b]$, corresponding to *puncture the ball*. Next, when activation decays following stimulation, it should land at a higher point for the representation having a higher current activation level than for the one having a lower current activation level (Sharwood Smith & Truscott, 2014, sec. 3.3.5). In other words, the resting activation level of the more active structure (*kick the ball*) should be higher than the resting activation level of the less active structure (*puncture the ball*). The structure with a higher resting activation level would be more readily available for subsequent processing.

1.3. The influence of lexical frequency on initial L2 syntax learning

The focus of this chapter is on whether and how lexical frequency facilitates adults' initial L2 syntax learning. The frequency with which words occur in the input has long been considered to play a role in L1 and L2 syntax acquisition. However, studies looking into this topic have focused on whether presenting a particular lexical item in a syntactic construction multiple times during exposure facilitates learning and generalization of this construction (e.g. L1, Casenhiser & Goldberg, 2005; Goldberg et al., 2004, 2007; L2, Denhovska et al., 2016; Mcdonough & Kim, 2009; Nakamura, 2012; Year & Gordon, 2009). For instance, Goldberg et al. (2004) had adult English natives learn a verb-final structure consisting of a subject NP, a locative NP and a nonsense verb ending in *–o* (*+ed*) expressing that the subject appeared at the position denoted by the second noun (e.g. *the rabbit the hat moopoed,* "the rabbit appeared in the hat"). They found that learning and generalisation of this structure was facilitated when a nonsense verb (e.g. *moopo*) was presented eight times in the exposure sentences and the other nonsense verbs were presented twice, as opposed to when verbs were presented either four times or twice during exposure. This and similar studies have been framed within usage-based models of first and second language acquisition, which maintain that acquisition is driven by experience with specific exemplars of the L1 or the L2 (e.g. Bybee, 2008; Goldberg, 2006; Tomasello, 2003). These models assume that children and adults are sensitive to the frequency with which certain combinations of lexical items occur in the language and that those patterns that occur more frequently are strengthened and become firmly established in the learner's mind. Accordingly, repeatedly presenting a construction with the same lexical item would facilitate learning of an item-specific construction. Then, as more exemplars with different lexical items are processed, the construction would be gradually generalized into an abstract syntactic representation.

To the best of my knowledge, no study has tested whether processing syntactic structures with high frequency vs. low frequency words (i.e. words differing in frequency of occurrence *in the natural language*) affects learning of these structures. In MOGUL, acquisition is the product of the processing activity, involving just processing mechanisms and differing from processing in no fundamental way (Sharwood Smith & Truscott, 2014, sec. 4.2). Considering this, I propose that if differences in lexical frequency affect syntax processing within this framework, as suggested in the previous section, these differences could also affect syntax acquisition by processing. In what follows, I hypothesize how differences in lexical frequency and, particularly, in verb frequency, would affect the initial acquisition of a cross-linguistically similar L2 structure (Section 1.3.1), a cross-linguistically dissimilar L2 structure (Section 1.3.2) and any learning differences between the two (Section 1.3.3). To exemplify this, I will focus on the cross-linguistically similar and dissimilar structures studied in this chapter (the same Spanish-Galician similar and dissimilar structures studied in Chapter 2). Crucially, for differences in lexical frequency to affect L2 syntax acquisition, learners must have been exposed to the L2 and must have encountered L2 words with different frequencies, so that some have a higher frequency of occurrence than others do. Nevertheless, I investigate syntax acquisition by learners who have never been exposed to the L2. I propose that, in this case, lexical frequency would be relevant as long as the lexical items containing the frequency manipulation are cognate in the L1 and in the L2 and the frequency manipulated is that of the words of the native language. For simplicity's sake, in Sections 1.3.1 to 1.3.3 I will discuss how lexical frequency could affect the acquisition of cross-linguistically similar and dissimilar structures from learners' first encounter with these constructions, but assuming that learners have processed the L2 words before, with a higher or a lower frequency. Then, in Section 1.3.4, I will comment on how the effects described could hold for complete beginner L2 learners. In my explanations, I will mostly concentrate on processing in the syntactic module. However, it must be remembered that, within MOGUL, sentence processing takes place simultaneously in all modules as words are incrementally perceived, and the processing system elaborates an acoustic and/or visual, phonological, syntactic and conceptual representation for each sentence (Sharwood Smith & Truscott, 2014, sec. 3.4.1).

1.3.1. The influence of lexical frequency on the acquisition of a cross-linguistically similar L2 structure

The Galician sentences in (2) are formed by a structure that exists in both Galician and Spanish. The embedded verb in (2a), *repare,* has a higher frequency of occurrence than the embedded verb in (2b), *desmonte* (frequency per million: *reparar,* 20 vs. *desmontar,* 8; Corpus de Referencia do Galego Actual, 2022). Let's consider how the two sentences would be processed when Spanish natives encounter them for the first time.

(2)	a.	É	importante	que	Pedro	repare	a	radio.
		it.is	important	COMP	Pedro	fix $_{\text{PRS.SBJV.3SG}}$	the	radio
		"It is important that Pedro fixes the radio."						
	b.	É	importante	que	Pedro	desmonte	a	radio.
		it.is	important	COMP	Pedro	disassemble $_{\text{PRS.SBJV.3SG}}$	the	radio
		"It is important that Pedro disassembles the radio."						

Upon reading or listening to *É importante,* the orthographic and phonological representations of these words would activate. Activation would spread in a bottom-up manner to the syntactic representations of the verb and the adjective (e.g. [V_j] and [ADJ_k]) and the syntactic processor would merge them into a VP (e.g. [$_{VP}$ V_j ADJ_k])[3]. Next, the VP would spread its activation to larger syntactic representations containing it. These representations would activate in parallel and would compete for selection as input is incrementally processed. The first time that the structure is encountered, the representations activated would be those acquired as part of the L1. If the input can be successfully processed using one of these, no new representation will be created; the L2 structure will simply be processed as if it was analogous to the L1 structure (Sharwood Smith & Truscott, 2014, sec. 4.2). One of the syntactic representations activated would be that of the cross-linguistically similar structure, possibly along the lines of [$_{CP}$ [$_{VP}$ V ADJ] [$_{CP}$ COMP [NP] [$_{VP}$ $V_{[+INFL]}$ [NP]]]][4] (as in Spanish *Es importante que Pedro repare la radio,* "It is important that Pedro fixes the radio"). Activation would weakly spread from [$_{CP}$ [$_{VP}$ V ADJ] [$_{CP}$ COMP [NP] [$_{VP}$ $V_{[+INFL]}$ [NP]]]] to its constituents. As this is occurring, the following words, *que Pedro repare a radio,* would be perceived. Processing of these words would involve the two streams of activation described. On the one hand, a

[3] Since all the words in these sentences are cognate between Spanish and Galician, upon reading or listening to the L2 items their L1 counterparts would activate too.

[4] $V_{[+INFL]}$ = finite verb.

bottom-up activation from the phonological representation of the words to coindexed syntactic representations, which would spread to larger representations containing them. This would gradually increase the activation of the cross-linguistically similar structure, causing it to dominate over competitors. On the other hand, a top-down activation from this structure to its components. The only phrase that could momentarily cause some processing difficulty for Spanish natives is *a radio,* because in Spanish *a* is a preposition usually followed by a noun (Bosque & Gutiérrez-Rexach, 2009). Yet, processing of the sentence so far (*É importante que Pedro repare*) would have strongly activated the representation $[_{CP}\ [_{VP}\ V\ ADJ][_{CP}\ COMP\ [NP]\ [_{VP}\ V_{[+INFL]}\ [NP]]]]$. Consequently, *a radio* would be processed as an NP, with the orthographic/phonological forms of *a and radio* coindexed with the syntactic representations of a determiner and a noun, respectively (e.g. $[_{NP}\ D_h\ N_s]$).

Crucially, since the embedded verb in (2a) is of a higher frequency than the embedded verb in (2b), the current activation level of the former will be higher than that of the latter. Anyhow, the activation of the verb's syntactic representation will spread to the syntactic representation of the structure containing it, raising, in turn, its current activation level. The activation spreading from the verb's syntactic representation to the representation of the cross-linguistically similar structure will be higher for the high frequency verb than for the low frequency verb. Consequently, the current activation level of the similar structure resulting from processing (2a) will be higher than the one resulting from processing (2b). When processing terminates, activation will fall back towards the resting level, landing at a position slightly above the original. The structure processed with a high frequency verb, having a high current activation level, should fall back to a higher resting activation level than the structure processed with a low frequency verb, having a lower current activation level (cf. Section 1.2). In this manner, each time that a sentence formed by the similar structure is processed, its resting activation level would increase, and the increases would be larger when the sentence contains a high frequency verb than a low frequency verb.

Importantly, MOGUL assumes that the relation between processing frequency and increases in resting activation level is logarithmic, i.e. the resting level increases rapidly when a representation is being developed and, then, the size of the increase gradually diminishes. At some point, a representation's resting activation level is so high that additional processing does not raise it any more (Sharwood Smith & Truscott, 2014, sec. 4.6.5)[5]. The cross-linguistically similar structure is assumed to be stored in the syntactic module at the beginning of L2 acquisition with a high resting activation level, as a result of its prior processing in the L1 (Sharwood Smith & Truscott, 2014, sec. 10.3). Thus, subsequent processing of this structure should result in small increases in its resting activation level (or possibly no increases, if its resting level were extremely high). Hence, it could be that the resting activation level of this

[5] This resonates with the relation between lexical frequency and lexical access speed presented in Section 1.1. Sharwood Smith and Truscott claim that their concept of resting activation level, which they apply to all the representations in the lexical stores, is identical to the concept described in models of lexical access. This is why they assume the relation between processing frequency and resting activation level to be logarithmic.

structure increases marginally more when processed with a high frequency verb compared to a low frequency verb. Yet, the overall difference in resting level derived from processing the structure with the two types of verbs may be small, or even imperceptible. This implies that the cross-linguistically similar structure may be comparably established in the linguistic system when processed with high frequency verbs and low frequency verbs.

1.3.2. The influence of lexical frequency on the acquisition of a cross-linguistically dissimilar L2 structure

The sentences in (3) are formed by a structure that exists only in Galician (it is dissimilar to the structure Spanish would use to express the same meaning). These contain the same embedded verbs as the sentences in (2). Thus, the verb in (3a) has a higher frequency than that in (3b). When Spanish natives encounter these sentences for the first time, they would process the first VP (*É importante*) as described for (2a) and (2b). This VP would spread its activation to other syntactic representations containing it. The first time that the structure is encountered, the representations activated would be those of the L1. These would include the representation of the cross-linguistically similar structure in the previous section.

(3) a. É importante Pedro reparar a radio.
it.is important Pedro fix$_{INF[AGR.3SG]}$ the radio
"It is important that Pedro fixes the radio."

b. É importante Pedro desmontar a radio.
it.is important Pedro disassemble$_{INF[AGR.3SG]}$ the radio
"It is important that Pedro disassembles the radio."

Next, the word *Pedro* would be processed and its syntactic representation (e.g. $[_{NP}\ N_t]$) would activate. At this point, the syntactic processor would detect a mismatch between the sentential input and any syntactic representations active in the syntactic module. A new representation (e.g. $[_{CP}\ [_{VP}\ V\ ADJ]\ [_{CP}\ NP]]$) would be created by selecting and combining the most active items in the syntactic store, simply for processing the input (Sharwood Smith & Truscott, 2014, sec. 4.2). Meanwhile, the verb (*reparar* or *desmontar*) would be perceived. Its syntactic representation (e.g. $[V_i]$) would activate and this activation would spread to its subcategorization frame (e.g. $[V_i\ NP_{p,\ q,\ r\ldots}]$). At the same time, activation would weakly spread from the verb's subcategorization frame to the syntactic representations of NPs and nouns previously used with the verb. As this is occurring, *a radio* would be processed, activating the syntactic representation of an NP (e.g. $[_{NP}\ D_h\ N_s]$, see previous section). The syntactic processor would merge all active representations, yielding a larger syntactic representation along the lines of $[_{CP}\ [_{VP}\ V\ ADJ][_{CP}\ [NP]\ [_{VP}\ V_{[-INFL]}\ [NP]]]$[6]. The representation created would receive a low resting activation level. When the structure is encountered again, the newly created syntactic

[6] $V_{[-INFL]}$ = non-finite (infinitive) verb.

representation would activate and the syntactic processor would select it amongst competitors.

Importantly, the current activation level of the embedded verb will be higher if it has a high frequency of occurrence, as in (3a), than if it has a low frequency of occurrence, as in (3b). In both cases, activation would spread from the syntactic representation of the verb to the representation of the cross-linguistically dissimilar structure. However, as described for the cross-linguistically similar structure, the higher activation of the high frequency verb compared to the low frequency verb would cause that the rise in the current activation level of the dissimilar structure is higher when including the former than the latter. In other words, the overall current activation level of the syntactic representation of (3a) should be higher than that of (3b). This would cause that when stimulation fades, the resting activation level of the first structure is higher than that of the second. That is, each time that a sentence formed by the cross-linguistically dissimilar structure is processed, its resting activation level would increase, and it would increase more when processed with a high frequency verb than with a low frequency verb. As mentioned, the relation between processing frequency and resting activation level is logarithmic. Since the initial resting activation level of the cross-linguistically dissimilar structure would be very low, the increases derived from additional processing should be quite large. Hence, the difference in resting level derived from processing the structure with high frequency vs. low frequency verbs should be significant. In brief, the cross-linguistically dissimilar structure should have a higher resting activation level and, thus, be more robustly established in the linguistic system, when processed with high frequency verbs than with low frequency verbs.

1.3.3. The influence of lexical frequency on the acquisition of a cross-linguistically similar vs. dissimilar L2 structure

As shown in Chapter 2, the fact that the cross-linguistically similar L2 structure can be processed using an L1 structure that is part of the linguistic system at the beginning of L2 acquisition, but the cross-linguistically dissimilar L2 structure has to be acquired from input, may lead to a learning advantage for the former over the latter. In MOGUL terms, this could be because the similar structure has a high resting activation level already at the start of L2 acquisition and this causes that, even if the dissimilar structure is learnt, the similar structure is more firmly established in learners' linguistic system. I propose that this should be the case regardless of whether the structures are processed with high frequency or low frequency verbs. On the other hand, the resting activation level of the cross-linguistically similar structure may be comparable as a result of processing sentences with high frequency and low frequency verbs (cf. Section 1.3.1), but the resting activation level of the cross-linguistically dissimilar structure may be higher as a result of processing sentences with high frequency verbs than with low frequency verbs (cf. Section 1.3.2). In familiar thinking terms, learning of the similar structure may be comparable with high frequency verbs and low frequency verbs, but learning of the dissimilar structure may be higher with high frequency verbs. If this is the case, then the difference in resting activation level between the similar and the dissimilar structure (i.e. the learning advantage of the former over the latter) may be smaller when sentences include high frequency verbs than low frequency verbs.

1.3.4. The influence of lexical frequency on L2 syntax learning by complete beginner learners

For differences in verb frequency to influence the acquisition of L2 structures as discussed, learners must have been exposed to L2 verbs with a higher or a lower frequency, so that when processing sentences formed by the target structures and containing a high or a low frequency verb, the difference in frequency is relevant. However, in this thesis I investigate syntax acquisition by complete beginner L2 learners, with their first exposure to the lexicon and the syntax of the L2 occurring in the lab. I argue that, in this case, the effects of verb frequency described would apply only if the L2 (Galician) verbs are cognate with L1 (Spanish) verbs of a higher or a lower frequency. In that case, when Spanish natives encounter a Galician cognate verb for the first time, its Spanish counterpart would activate. To process the L2 verb, its orthographic and phonological representations would have to be coindexed with a syntactic and a conceptual representation (Sharwood Smith & Truscott, 2014, sec. 4.2). These would be the representations of its L1 counterpart. Crucially, by the processes described in Section 1.1, the syntactic and conceptual representations of L1 high frequency verbs would be more strongly activated (i.e. would have a higher current activation level) than those of L1 low frequency verbs. In MOGUL, activation flows from one module to another in a bottom-up and a top-down manner, as well as within modules (Sharwood Smith & Truscott, 2014, sec. 3.4.1). Thus, I assume that the higher or lower activation of the syntactic and the conceptual representations of L1 verbs would spread to coindexed phonological and orthographic representations in the L1 and the L2. Similarly, the activation of the phonological representation of L1 high and low frequency verbs would spread to the phonological representation of L2 verbs. In both cases, activation would be stronger when coming from a high frequency verb than from a low frequency verb. In sum, L2 cognates of L1 high frequency verbs would be more strongly activated than L2 cognates of L1 low frequency verbs. This stronger activation would then spread to the cross-linguistically similar or dissimilar structure containing them, affecting acquisition by processing as described. This resonates with evidence that when learners process cognates and interlingual homographs in the L2, the representations of these words in the L1 and the L2 activate and learners are sensitive to the frequency of words in the two languages (e.g. Dijkstra et al., 1998).

2. Experiment 3

2.1. Overview

To investigate whether differences in lexical frequency affected L2 syntax acquisition as described, I conducted Experiment 3, which was designed to be compared with Experiment 2. Experiment 3 exposed Spanish natives with no knowledge of Galician to the Spanish-Galician similar and dissimilar structures studied in Experiment 2 and examined how well the two structures were established in learners' linguistic system. The only difference between

Experiment 2 and Experiment 3 was a lexical frequency manipulation: the cognate verbs in the experimental sentences of Experiment 3 had a significantly lower mean frequency of occurrence in Spanish than the cognate verbs in Experiment 2. Hence, the research question I addressed in Experiments 2 vs. 3 was *"Do differences in lexical frequency facilitate initial L2 syntax acquisition? If so, is this facilitation modulated by the cross-linguistic similarity of structures between the L1 and the L2?".* The mini-language used in Experiment 2 was also used in Experiment 3, with the exception of verbs, which, as mentioned, changed from the first experiment to the second. The procedure of Experiment 3 was almost identical to that of Experiment 2. First, learners were exposed to the cross-linguistically similar and dissimilar structures in a structure-search task. The exposure phase was followed by a testing phase consisting of a Grammaticality Judgement Task (GJT) with feedback. Then, learners were encouraged to express any knowledge about the structures in a verbal report. The experiment ended with a vocabulary test to make sure that participants were familiar with the verbs in which the lexical frequency manipulation took place.

My hypotheses were advanced in Section 1.3 and can be summarised as follows. **Hypothesis 1 (H1)** claimed that the similar structure would be comparably established in the linguistic system as a result of processing sentences with high frequency verbs (Experiment 2) and low frequency verbs (Experiment 3). In familiar thinking terms, I hypothesized that lexical frequency would not facilitate the acquisition of the similar structure. **Hypothesis 2 (H2)** claimed that the dissimilar structure would be more robustly established in the linguistic system as a result of processing sentences with high frequency verbs (Experiment 2) than low frequency verbs (Experiment 3). In other words, I hypothesized that lexical frequency would facilitate the acquisition of the dissimilar structure. Finally, **Hypothesis 3 (H3)** claimed that the similar structure would be more robustly established in the linguistic system than the dissimilar structure as a result of processing sentences with high frequency verbs and low frequency verbs. However, the difference between the two structures would be smaller when processed with high frequency verbs (Experiment 2) than with low frequency verbs (Experiment 3). Put differently, I hypothesized that there would be a learning advantage for the similar structure over the dissimilar structure, but that this advantage would be smaller when the structures were processed with high frequency verbs than when they were processed with low frequency verbs.

2.2. Participants

The number of participants and the participant profile in Experiment 3 was the same as in Experiment 2 (see Chapter 2, Section 3.2 for a description of participants in Experiment 2). Forty-four native speakers of Spanish (34 female) took part in Experiment 3. They were aged between 19 and 33 (M = 22, SD = 2.97) and they were all students at the University of the Basque Country (UPV/EHU). None of them had knowledge of Galician or other languages with inflected infinitives, as reported in the linguistic background questionnaire filled out before the experiment (the same as in Experiment 2, see Appendix A-1). Most participants (75%, 33/44) had at least some knowledge of Basque and their proficiency level in this language was at B2 or under. However, they all claimed that they felt the most comfortable using Spanish. The 44 participants were spoken to only in Spanish prior to entering school (0-3 years).

The linguistic background questionnaire asked about the average frequency with which participants used Spanish and Basque during their childhood (3-12 years), puberty (12-18 years) and adulthood (after 18 years) at school/university/work, at home or at other places. Responses ranged from 1 (*Spanish only*) to 7 (*Basque only*). In short, like in Experiment 2, in Experiment 3 Spanish was the language participants currently used and had used the most throughout their lives (mean language use during childhood, Experiment 2, 1.52 (*SD* = 1.14), Experiment 3, 1.76 (*SD* = 1.24); puberty, Experiment 2, 1.54 (*SD* = 1.16), Experiment 3, 1.77 (*SD* = 1.23); adulthood, Experiment 2, 1.24 (*SD* = 0.54), Experiment 3, 1.32 (*SD* = 0.54)). The mean frequency of use of Spanish did not significantly differ between experiments, as compared by independent-samples t-tests (all $p \geq .10$, see Appendix B-1 for more details, including the mean scores reported by participants in Experiments 2 and 3 for each environment in each period of life). In addition, the linguistic background questionnaire asked about self-assessed proficiency level speaking, listening, reading and writing in Spanish. Responses also ranged from 1 (*very poor proficiency*) to 7 (*perfect proficiency*). In sum, like in Experiment 2, participants in Experiment 3 reported having a nearly perfect mastery of Spanish, and this was comparable across experiments (mean proficiency level collapsing all four skills, Experiment 2, 6.77 (*SD* = 0.51) vs. Experiment 3, 6.81 (*SD* = 0.39), t (350) = -0.82, p = .41, d = -0.09). The mean scores for each skill reported by participants in Experiments 2 and 3 are available in Appendix B-1.

All participants had normal or corrected to normal vision and hearing. Prior to the experiment, they read and signed an informed consent (see Appendix B-2). Experiment 3 was part of the project "Cross-linguistic influence in language learning, processing and aging" (PID2021-124056NB-I00), funded by the Spanish Ministry of Science, Innovation and Universities and approved by the Committee of Ethics for research involving human beings of the University of the Basque Country (*Comité de Ética para las Investigaciones con Seres Humanos, CEISH,* Ref. M10_2022_317). Participants received 7€ for their participation.

2.3. Materials

As advanced, the mini-language used in Experiment 2, consisting of a Spanish-Galician cognate vocabulary and Galician-based syntax was used in Experiment 3, with some vocabulary modifications. The object of study of Experiment 3 were the two structures varying in cross-linguistic similarity between Spanish and Galician labelled as *similar structure* and *dissimilar structure* in Chapter 2. The two constructions were described in detail in that chapter (Section 2.3) and were exemplified again in this chapter (Section 1.3). Here, I summarise their main characteristics for ease of reading. Both the similar and the dissimilar structure were embedded SVO clauses following an impersonal expression conveying opinion (e.g. *É importante,* "it is important") and expressed the same meaning. While the similar structure existed in Spanish and Galician, the dissimilar structure was adapted from a construction existing only in Galician (the normally post-verbal subject was made pre-verbal to make the structure more comprehensible for Spanish natives, see Chapter 2, Section 1.2). The main difference between the structures was that, while the embedded clause in the similar structure was headed by the complementizer *que* ("that") and contained a verb in the present subjunctive, the embedded clause in the dissimilar structure was not headed by *que* and contained an inflected infinitive. In both embedded clauses, a third person singular subject agreed with the verb, so the inflected infinitive took a covert person and number mark and was formally identical to the non-inflected infinitive (Real Academia Galega & Instituto da Lingua Galega, 2012). Table 3.1 displays example sentences formed by the two structures taken from Experiment 3.

Structure	Example
Similar structure	É importante [que Pedro desmonte a radio] it.is important COMP Pedro disassemble $_{PRS.SBJV.3SG}$ the radio "It is important that Pedro disassembles the radio."
Dissimilar structure	É importante [Pedro desmontar a radio] it.is important Pedro disassemble $_{INF[AGR.3SG]}$ the radio "It is important that Pedro disassembles the radio."

TABLE 3.1. Example sentences formed by the similar and the dissimilar structure taken from Experiment 3. Embedded clauses are between brackets.

Using these structures, I generated the sentences displayed in the exposure phase and the testing phase (the *exposure set* and the *testing set*, the latter also including ungrammatical sentences). As mentioned, sentences in Experiment 2 and Experiment 3 were almost identical, except for the embedded verbs in the exposure and the testing sets. In Experiment 3, these were selected so that, in all cases, the lemmatic frequency per million of their Spanish translations was lower than in Experiment 2 as measured in the Corpus of the 21st century Spanish (CORPES XXI, Real Academia Española, 2021)[7]. Sentences in both the exposure and the testing sets of Experiment 3 contained verbs of a lower frequency than in Experiment 2 because, since the GJT included feedback, I assumed that some learning could take place during the testing phase (this was already reported to occur in Experiment 2, see Chapter 2, Section 3.7.1 for more details).

Like in Experiment 2, the sentences in Experiment 3 were created with the online automatic translator *Gaio* from the General Secretary for language policy of the Galician government (*Secretaría xeral de política lingüística*, *Xunta de Galicia*). A Galician native speaker

[7] The CORPES XXI was considered a reliable data set of frequencies of occurrence in Spanish because it includes a large number of words (350 million orthographic forms annotated with syntactic category information) and is based on a large sample of written and oral transcriptions (over 327,000 documents, comprising novels, books, theatre plays, magazines, newspapers, blogs, interviews and more). Additionally, the corpus is regularly updated. Frequencies were looked up in the version published in 2021.

corroborated that the vocabulary, verb conjugation and article use was accurate. Because, just as in Experiment 2, participants would be exposed to sentences aurally (and in written form) during the exposure and testing phases, the same female native speaker of Galician that recorded the exposure set and the testing set of Experiment 2 recorded the exposure set and the testing set of Experiment 3. Like in Experiment 2, sentences were recorded one by one in a soundproof booth with an Olympus voice recorder (Linear PCM Recorder LS-5 model, frequency sampling of 96kHz), at a normal pace and with natural intonation. The initial and final silences in each recording were cut using *Praat* (Boersma & Weenink, 2018, version 6.0.37).

2.3.1. Exposure set

The exposure set of Experiment 3 was based on the exposure set of Experiment 2. Both consisted of 100 grammatical sentences formed by the similar structure and the corresponding 100 grammatical sentences formed by the dissimilar structure. However, in Experiment 3 the exposure verbs in Experiment 2 were replaced by novel Spanish-Galician cognate verbs according to a frequency criterion. As advanced, the verbs in Experiment 3 were selected so that the lemmatic frequency per million of their Spanish translations in the CORPES XXI was below the lowest frequency value of the Spanish translations of verbs in Experiment 2. Sentences were divided into two lists, so that participants were not exposed to the two versions of a sentence. Each list included 50 sentences formed by the similar structure and 50 formed by the dissimilar structure. The lexicon of the exposure set was identical to that of Experiment 2, except for 20 novel embedded verbs (to be consulted in Appendix B-3). Apart from this, the lexicon consisted of the same five impersonal expressions conveying opinion, the same 20 proper nouns functioning as the subject of the embedded clause and the same 20 inanimate nouns functioning as the direct object of the embedded verb as in Experiment 2 (reported in Appendix A-4). Exposure verbs in Experiments 2 and 3 were matched in length (number of letters) and degree of cognateness, as assessed by the phonological and orthographic overlap between the Galician verbs and their Spanish counterparts (Levenshtein distance)[8]. Importantly, they differed in the mean lemmatic frequency per million of their Spanish translations, as indicated by an independent-samples t-test (Experiment 2, M = 98.97 (SD = 106.16) vs. Experiment 3, M = 4.76 (SD = 2.02), t (38) = 3.96, p < .001, large effect size of d = 1.25). As in Experiment 2, each impersonal expression occurred 10 times in sentences formed by the similar structure and 10 times in sentences formed by the dissimilar structure per list. Proper nouns, verbs and inanimate nouns occurred twice or three times per condition across lists. These lexical items did not statistically differ in terms of length (number of letters), frequency per million of their Spanish counterparts (in the CORPES XXI) and level of phonological and orthographic overlap with their Spanish translations (Levenshtein distance) between conditions in the two lists (all p ≥ .38, see Appendix B-4). Each impersonal expression

[8] Independent-samples t-tests comparing verb length: t (38) = -1.92, p > .05, phonological and orthographic Levenshtein distance: t (38) = -1.45, p = .15 and t (38) = 1.29, p = .20, respectively.

occurred with a specific proper noun, verb and inanimate noun only once in the exposure set. This can be found in Appendix B-5.

2.3.2. Testing set

In the testing phase, participants conducted a GJT. Hence, the testing set consisted of grammatical and ungrammatical sentences. The testing set of Experiment 3 was identical to that of Experiment 2 except for the verbs in the embedded clauses, which were chosen following the same frequency criterion described in the previous section. The testing set included, thus, 80 grammatical sentences formed by the similar structure and the corresponding 80 grammatical sentences formed by the dissimilar structure, as well as 80 ungrammatical sentences violating the similar structure and 80 ungrammatical sentences violating the dissimilar structure derived from the grammatical ones. As a reminder, sentences violating the similar structure contained an embedded verb in the infinitive and sentences violating the dissimilar structure contained an embedded verb in the present subjunctive (see Table 3.2 for examples of the four conditions). The four versions of a given sentence occurred in four different lists. Each list contained 80 items (20 grammatical sentences formed by the similar structure and 20 ungrammatical sentences violating it, 20 grammatical sentences formed by the dissimilar structure and 20 ungrammatical sentences violating it).

Condition	**Example**[9]
Grammatical similar structure (SS)	É importante que Antonio transcriba a carta it.is important COMP Antonio transcribe $_{\text{PRS.SBJV.3SG}}$ the letter "It is important that Antonio transcribes the letter."
Ungrammatical similar structure (*SS)	*É importante que Antonio transcribir a carta it.is important COMP Antonio transcribe $_{\text{INF[AGR.3SG]}}$ the letter "It is important that Antonio transcribes the letter."
Grammatical dissimilar structure (DS)	É importante Antonio transcribir a carta it.is important Antonio transcribe $_{\text{INF[AGR.3SG]}}$ the letter "It is important that Antonio transcribes the letter."
Ungrammatical dissimilar structure (*DS)	*É importante Antonio transcriba a carta it.is important Antonio transcribe $_{\text{PRS.SBJV.3SG}}$ the letter "It is important that Antonio transcribes the letter."

TABLE 3.2. Examples of the four experimental conditions of the testing set in Experiment 3. Following the convention in linguistics, ungrammaticality is indicated by an asterisk.

The lexicon of the testing set was shared with Experiment 2, with the exception of verbs (available in Appendix B-3). Hence, it consisted of the five impersonal expressions in the exposure set, and 20 verbs and 40 nouns (20 proper nouns —the subject of the embedded clause—and 20 inanimate nouns —the direct object of the clause) different from the ones in

[9] I provide the intended English translation for all sentences irrespective of their grammaticality.

the exposure set (reported in Appendix A-4 for Experiment 2). Independent-samples t-tests indicated that the verbs in the testing sets of Experiments 2 and 3 were similar in number of letters and cognateness, i.e. phonological and orthographic Levenshtein distance with their Spanish translations[10]. Crucially, the mean lemmatic frequency per million of the Spanish translations of verbs in Experiment 3 (M = 5.13, SD = 2.55) was significantly lower than in Experiment 2 (M = 115.09, SD = 112.98), as indicated by an independent-samples t-test (t (38) = 4.35, p < .001, large effect size of d = 1.38). As in Experiment 2, each impersonal expression occurred four times per condition across lists. Proper nouns, inanimate nouns and verbs occurred once per condition across lists. An impersonal expression, proper noun, verb and inanimate noun occurred together only once in the testing set. This can be consulted in Appendix B-5.

2.4. Procedure

The procedure of Experiment 3 was virtually identical to the one of Experiment 2. Participants were told that they would participate in a study about sentence comprehension in Galician. The experiment included three phases: an exposure phase, consisting of a structure-search task (Section 2.4.1), a testing phase, consisting of a GJT (Section 2.4.2) and a debriefing phase, consisting of a verbal report and a vocabulary test (Section 2.4.3). Mirroring Experiment 2, all tasks were run on the Eprime 2.0 software. Each participant was tested individually in a soundproof booth. The duration of the experiment was around 45 minutes. Instructions were given in Spanish and were identical to the ones in Experiment 2 (cf. Appendix A-10).

2.4.1. Exposure phase

The procedure of the exposure phase was like that in Experiment 2 (see Chapter 2, Section 3.4.1). In brief, participants were aurally and visually presented with sentences formed by the similar structure and the dissimilar structure and were instructed to pay attention to the form of the sentences to identify the two structures according to which they were formed. They were informed that they would be tested on these structures in the following part of the experiment. Each participant was presented with one of the two lists of sentences in the exposure set, one by one and in a randomized order. Participants listened to their exposure list twice, so they were actually exposed to 200 sentences (100 formed by the similar structure and 100 formed by the dissimilar structure). Each sentence appeared written at the centre of a white computer screen and it was simultaneously played through headphones. Sentences were automatically presented one after the other. The duration of the exposure phase was around 10 minutes.

[10] Independent-samples t-tests comparing verb length: t (38) = -1.57, p = .13, phonological and orthographic Levenshtein distance: t (38) = -0.42, p = .68 and t (38) = 0, p = 1, respectively.

2.4.2. Testing phase

The procedure of the testing phase was also identical to the one in Experiment 2 (see Chapter 2, Section 3.4.2). In short, participants were told that they would read and listen to novel sentences in Galician. They were instructed to decide, as quickly as possible, whether these sentences were formed by one of the two structures that they had read and listened to in the previous part of the experiment or not. Each sentence had to be judged as "correct" or "incorrect". Sentences were presented one by one and in a randomized order for each participant. Each sentence was played and, at the same time, it was written at the centre of a white computer screen. When a stimulus ended, the options *Correcto* ("Correct") and *Incorrecto* ("Incorrect") replaced the sentence. Participants pressed the key "A" if they thought a sentence was correct and the key "L" if they thought it was incorrect. If no key was pressed within five seconds, the next sentence was automatically displayed. If participants pressed a key, they saw a green tick (if their answer was right) or a red cross (if their answer was wrong) for 700ms. Before the test, a brief practice session with a sentence of each condition not included in the testing set (a grammatical similar structure, a grammatical dissimilar structure, an ungrammatical similar structure and an ungrammatical dissimilar structure) helped participants familiarize themselves with the task.

2.4.3. Debriefing phase

2.4.3.1. Verbal report

The verbal report was also like in Experiment 2 (Chapter 2, Section 3.4.3). Participants were reminded that they had been exposed to "two structures" and they were encouraged to put into words any knowledge about them. Specifically, they were asked:

1. As mentioned during the experiment, the sentences you listened to were formed according to two structures. Did you notice which structures were these?
2. If yes, please indicate which you think these structures are.

Participants typed their answers in a text box. Since learning occurred under intentional conditions (i.e. with learners consciously trying to deduce the structures under study), participants were expected to develop conscious knowledge of the two structures and to be able to provide metalinguistic descriptions of them. Expected descriptions were:

- **Similar structure:** A structure contains the word *que* and then a verb conjugated (in the present subjunctive).
- **Dissimilar structure:** A structure does not contain the word *que* and contains a verb in the infinitive.

2.4.3.2. Vocabulary test

The experiment ended with a vocabulary test evaluating participants' familiarity with the Spanish translations of the Galician cognate verbs used in the exposure and testing sets (see Appendix B-6). For the frequency manipulation in this study to be effective, it was important that participants in the two experiments compared (Experiment 2 and Experiment 3) were familiar with the Spanish equivalents of the Galician verbs. This is because, as discussed in Section 1.3.4, I assumed that upon reading and listening to a cognate verb, the orthographic, phonological, syntactic and conceptual representations of its Spanish counterpart would activate, with activation being higher for Spanish high frequency verbs than for Spanish low frequency verbs. I assumed that the higher or lower activation of the Spanish verbs would translate into the Galician verbs (i.e. Galician verbs cognate with Spanish high frequency verbs would be more strongly activated than Galician verbs cognate with Spanish low frequency verbs). This activation would spread from the syntactic representation of the verb to the syntactic representation of the similar or dissimilar structure containing it, ultimately affecting its acquisition by processing. Hence, if participants were not familiar with the Spanish counterparts of the Galician verbs, when processing the Galician cognate verbs the equivalent Spanish verbs would not activate and the frequency manipulation effectuated would not be adequate.

On the one hand, in Experiment 3, participants read and listened to sentences containing Spanish-Galician cognate verbs with a low lemmatic frequency per million in Spanish (cf. Section 2.3). This means that, according to the written and oral texts in the CORPES XXI, these verbs do not occur very often in the language. Therefore, it could be that some of them were unknown to participants. Because of this, participants conducted the vocabulary test, administered in pen and paper format. On the other hand, I assumed that participants in Experiment 2 were familiar with the Spanish counterparts of the Galician verbs in their experiment, since these had a much higher mean frequency of occurrence than the verbs in Experiment 3 (cf. Section 2.3). Additionally, they were pretty common verbs (e.g. sell, buy, win, lose, cut). Nevertheless, to make sure that these participants were familiar with the verbs used in Experiment 2, they were contacted by email and were asked to conduct an online version of the vocabulary test with the verbs in the exposure and the testing sets in Experiment 2. Anyhow, in the vocabulary test participants were given a list of the Spanish translations of the Galician verbs in the infinitive form. They had to indicate whether they knew the meaning of each verb putting a tick next to it. Participants could be familiar with a verb but still report not knowing its meaning, for instance, because they were unsure about what this meaning was or because they did not have a clear, concrete semantic representation of it. For this reason, if they reported not knowing the meaning of a verb, they were asked two follow-up questions:

1. Have you heard this verb before? (yes/no)
2. Could you interpret it in the context of a sentence? (yes/no)

2.5. Predictions

In Section 2.1, I summarised my three hypotheses regarding how exposure to the similar and the dissimilar structure with high frequency verbs (Experiment 2) vs. low frequency verbs (Experiment 3) would affect the establishment of the two structures in the linguistic system. Whether these hypotheses are correct or not will be shown by a comparison of the performance in the GJT of Experiments 2 and 3. Like in Experiment 2, in Experiment 3 accuracy and *d'* analyses evaluated performance in this task. I assumed that learning of a structure would be indicated by learners' capacity of accepting that structure and rejecting its ungrammatical equivalent (accuracy analysis) or by learners' sensitivity to the difference between a grammatical structure and its ungrammatical counterpart (*d'* analysis).

Hypothesis 1 (H1) claimed that the similar structure would be comparably established in the linguistic system as a result of processing sentences with high frequency verbs (Experiment 2) and low frequency verbs (Experiment 3). In familiar thinking terms, I hypothesized that lexical frequency would not facilitate the acquisition of the similar structure. **If H1 was correct,** I predicted, first, that like in Experiment 2, in Experiment 3 the GJT would evidence that the similar structure was learnt. That is, I predicted that learners would judge the grammatical similar structure and the ungrammatical similar one significantly above chance. Additionally or alternatively, learners would be sensitive to the difference between these two structures. Second, if the establishment of the similar structure in learners' linguistic system was comparable when processed with high frequency verbs and with low frequency verbs, I predicted that accuracy for the similar structure and its ungrammatical counterpart and/or sensitivity to the difference between the two would be comparable in Experiments 2 and 3.

Hypothesis 2 (H2) claimed that the dissimilar structure would be more robustly established in the linguistic system as a result of processing sentences with high frequency verbs (Experiment 2) than with low frequency verbs (Experiment 3). In other words, I hypothesized that lexical frequency would facilitate the acquisition of the dissimilar structure. **If H2 was correct,** I predicted, first, that like in Experiment 2, in Experiment 3 the test would show that the dissimilar structure was learnt. That is, I predicted that learners would judge the grammatical dissimilar structure and the ungrammatical dissimilar one significantly above chance. Additionally or alternatively, learners would be sensitive to the difference between these two structures. Second, if learning of the dissimilar structure was greater when processed with high frequency verbs than with low frequency verbs, I predicted that accuracy for the dissimilar structure and its ungrammatical counterpart and/or sensitivity to the difference between the two would be significantly higher in Experiment 2 than in Experiment 3.

Finally, **Hypothesis 3 (H3)** claimed that the similar structure would be more robustly established in the linguistic system than the dissimilar structure as a result of processing sentences with high frequency and low frequency verbs. Yet, the difference between the two structures would be smaller when processed with high frequency verbs (Experiment 2) than with low frequency verbs (Experiment 3). **If H3 was correct,** I predicted, first, that like in

Experiment 2, in Experiment 3 accuracy would be significantly higher when judging the grammatical and the ungrammatical similar structure than the grammatical and the ungrammatical dissimilar structure. Additionally or alternatively, sensitivity to the difference between the first two structures would be significantly higher than to the difference between the last two structures. Second, I predicted that the magnitude and/or the strength of the difference in accuracy and *d'* scores when judging the grammatical and the ungrammatical similar structure vs. the grammatical and the ungrammatical dissimilar ones would be smaller in Experiment 2 than in Experiment 3. That is, learners in Experiment 2 would obtain smaller effect sizes and/or larger p-values than learners in Experiment 3 in the tests comparing accuracy and *d'* scores for the two pairs of structures.

2.6. Coding and data analysis

In this section, I present how I analysed Experiment 3, comparing it when necessary with Experiment 2. Experiment 3 was analysed using the programming environment R (R Core Team, 2022, version 4.2.2). The function and package with which each statistical test and effect size were computed are reported just the first time that a test or effect size is stated.

2.6.1. Testing phase

Responses in the GJT were coded as binary (1 = correct grammaticality judgement, 0 = incorrect grammaticality judgement). Trials in which participants did not provide a response were removed from the analysis, for E-prime coded accuracy in these trials as 0 when actually no judgement was made. This corresponded to 0.34% (3/880) of all grammatical similar structures, 0.11% (1/880) of all ungrammatical similar structures and 0.11% (1/880) of all ungrammatical dissimilar structures.

The similar and the dissimilar structure as part of the linguistic system

I started by evaluating whether the similar and the dissimilar structure were established in learners' linguistic system (for learners in Experiment 2, Chapter 2 already showed this to be true). I looked at mean accuracy percentages when judging the grammatical similar structure, the grammatical dissimilar structure, the ungrammatical similar structure and the ungrammatical dissimilar structure. Then, I compared each percentage against chance (50%) using one-sample t-tests fitted with the function *t.test* from the *stats* package (R Core Team, 2022). Next, I evaluated whether the accuracy analysis could be influenced by participants having a bias towards accepting or rejecting the structures. To this aim, responses in the test were coded as Hits, False alarms, Misses or Correct rejections. Then, I computed the mean index of response bias *c* from the Signal Detection Theory for (i) the grammatical and the ungrammatical similar structure and (ii) the grammatical and the ungrammatical dissimilar structure. These indices were compared against zero (indicating no response bias) by means of one-sample t-tests. *Cohen's d* was the standardised measure of effect size for these t-tests. This was calculated using the function *cohens_d* from the *rstatix* package (Kassambara, 2021).

A *d* of 0.2, 0.5 and 0.8 was considered small, medium and large, respectively (Cohen, 1988)[11]. Because the analysis revealed a significant response bias, I performed a *d′* analysis to determine whether participants were sensitive to the difference between the grammatical and the ungrammatical similar structure, on the one hand, and the grammatical and the ungrammatical dissimilar structure, on the other hand, irrespective of response bias. I calculated *d′* scores for each participant using the function *dprime* from the *psycho* package (Makowski, 2018). Then, I compared mean *d′* scores against chance (zero) using one-sample t-tests. As will be detailed in Section 2.7.1, both the accuracy and the *d′* analyses indicated that, like in Experiment 2, in Experiment 3 the similar and the dissimilar structure were part of learners' linguistic system. In this light, I turned to examining whether differences in lexical frequency affected the establishment of the structures as predicted by my hypotheses, comparing the performance of participants in Experiments 2 and 3.

Comparing the establishment of the similar and the dissimilar structure in the linguistic system

If the predictions of my hypotheses were correct, verb frequency (high frequency, Experiment 2 and low frequency, Experiment 3) should have affected differently accuracy when judging the grammatical and the ungrammatical similar structure, and the grammatical and the ungrammatical dissimilar structure. Mirroring the analyses in Chapter 2, I collapsed accuracy for the grammatical and the ungrammatical similar structure, on the one hand, and the grammatical and the ungrammatical dissimilar structure, on the other hand. Then, a generalized linear mixed effects model fitted with the function *glmer* from the *lme4* package (D. Bates, Mächler, et al., 2015) tested for the interaction between the effect of Cross-linguistic similarity (Similar structure vs. Dissimilar structure) and Verb frequency (High vs. Low) on accuracy. The model that converged included random intercepts by participant and by item. I used deviation coding for the variables Cross-linguistic similarity and Verb frequency. The categories *Similar structure* and *High* were assigned the value 0.5 and the categories *Dissimilar structure* and *Low* were assigned the value -0.5[12].

Because the test yielded a significant interaction (see Section 2.7.1), I conducted post-hoc pairwise comparisons by verb frequency and cross-linguistic similarity. On the one hand, separate generalized linear mixed effects models tested for the effect of Verb frequency on accuracy when judging (i) the similar structure and its ungrammatical counterpart and (ii) the dissimilar structure and its ungrammatical counterpart. These models included by-participant and by-item random intercepts. Treatment coding was used for the variable Verb frequency, so that the category *Low* received the value 0 and the category *High,* the value 1. On the other hand, generalized linear mixed effects models assessed the effect of Cross-linguistic similarity on accuracy for (i) participants in Experiment 2, who processed sentences with high frequency

[11] This is the effect size for all t-tests in this chapter.

[12] Remember that this coding scheme facilitates the interpretation of the effects of the independent variables on the dependent variable as main effects in the presence of interactions (Sonderegger et al., 2018).

verbs, and (ii) participants in Experiment 3, who processed sentences with low frequency verbs. The models fitted had random intercepts by participant and by item and a by-participant random slope of Cross-linguistic similarity to account for the fact that the effect of this variable on accuracy could vary for each participant. Treatment coding was used for the variable Cross-linguistic similarity, so that the category *Dissimilar structure* received the value 0 and the category *Similar structure,* the value 1.

Turning to the *d'* analysis, I computed two *d'* scores for each participant in Experiments 2 and 3, one indicating sensitivity to the difference between the grammatical and the ungrammatical similar structure and the other indicating sensitivity to the difference between the grammatical and the ungrammatical dissimilar structure. To test whether verb frequency affected *d'* scores for each of these pairs of structures differently, I ran a 2x2 ANOVA with *d'* scores as the dependent variable and the interaction between two independent variables: Cross-linguistic similarity (coded as a within-subjects variable) and Verb frequency (coded as a between-subjects variable). The ANOVA was fitted with the function *anova_test* from the *rstatix* package. As a standardized effect size measure for the results of the test, I calculated partial eta-squared (η_p^2). A η_p^2 of 0.01, 0.06 and 0.14 was considered small, medium and large, respectively (Cohen, 1969). The test yielded a significant interaction (cf. Section 2.7.1 below) so, as in the accuracy analysis, I conducted post-hoc pairwise comparisons by verb frequency and cross-linguistic similarity.

On the one hand, two independent-samples t-tests tested for differences in sensitivity as a function of verb frequency when judging (i) the similar structure and its ungrammatical counterpart and (ii) the dissimilar structure and its ungrammatical counterpart. A series of Shapiro-Wilk tests indicated that the sample of *d'* scores reflecting sensitivity to the difference between the similar structure and the structure violating it was not normally distributed nor for learners in Experiment 2 ($W = 0.85$, $p < .001$) nor for learners in Experiment 3 ($W = 0.79$, $p <.001$). Likewise, the sample of *d'* scores reflecting sensitivity to the difference between the dissimilar structure and the structure violating it was not normally distributed nor for learners in Experiment 2 ($W = 0.89$, $p <.001$) nor for learners in Experiment 3 ($W = 0.84$, $p <.001$). Nevertheless, because in all cases the sample size was greater than 30, this was compatible with the parametric independent-samples t-tests (Levshina, 2015). Levene's tests revealed that the variances of the samples to be compared were equal. Specifically, when comparing the *d'* scores for the difference between the similar structure and its ungrammatical counterpart in Experiment 2 vs. 3, the result of the test was F (1, 86) = 1.28, $p = .26$. When comparing the *d'* scores for the difference between the dissimilar structure and its ungrammatical counterpart in Experiment 2 vs. 3, the output of the test was F (1, 86) = 0.43, $p = .51$.

On the other hand, two paired-samples t-tests assessed the effect of cross-linguistic similarity on *d'* scores. Specifically, the tests compared the *d'* scores reflecting sensitivity to the difference between the grammatical and the ungrammatical similar structure vs. the *d'* scores reflecting sensitivity to the difference between the grammatical and the ungrammatical dissimilar structure (i) in Experiment 2 (with high frequency verbs) and (ii) in Experiment 3 (with low frequency verbs). In this case, Shapiro-Wilk tests showed that the differences between the pairs of *d′* scores were not normally distributed nor in the first experiment (W = 0.91, p = .003) nor in the second (W = 0.94, p = .02). However, once again the sample size was larger than 30, so I computed the parametric paired-samples t-tests. Levene's tests revealed that the variances in the populations that represented the pairs of *d′* scores in Experiments 2 and 3 were homogeneous (F (1, 86) = 0.23, p = .63 and F (1, 86) = 2e-4, p = .99, respectively). All Shapiro-Wilk tests were computed with the function *shapiro.test* from the *stats* package (R Core Team, 2022) and all Levene's tests were computed with the function *leveneTest* from the *car* package (Fox & Weisberg, 2019)[13]. All figures illustrating the results of this study were made using the function *ggplot* from the *ggplot2* package (Wickham, 2016).

2.6.2. Debriefing phase

2.6.2.1. Verbal report

Participants' responses in the verbal report were transcribed. The same two experimenters who classified participants in Experiment 2 as *aware* or *unaware* of the similar and the dissimilar structure assessed awareness for participants in Experiment 3, and using the same rubric (the transcriptions and the rubric are available in Appendix B-7, see Appendix A-8 for the transcription of the verbal reports in Experiment 2). As a reminder, a participant was considered *aware* if s/he could report one or the two structures (see Section 2.4.3 for expected descriptions). A participant was considered *unaware* if s/he could not identify that the two structures differed in the presence/absence of the complementizer *que* and the finite or non-finite nature of the embedded verb. Alternatively, *unaware* participants could identify the varying elements in the structures but could not correlate them as required. Disagreements were discussed until reaching a unanimous decision. Like in Experiment 2, awareness results were coded as binary (1 = aware participant, 0 = unaware participant) and I calculated the percentage of aware and unaware participants.

2.6.2.2. Vocabulary test

I analysed the results of the vocabulary test conducted by participants in Experiment 3, as well as the results of the online version of the test conducted by participants in Experiment 2. As a reminder, participants were first asked whether they knew the meaning of the Spanish

[13] To corroborate that the parametric tests were not affected by the non-normality of the samples of *d′* scores compared (independent-samples t-tests) or the differences between the pairs of *d′* scores compared (paired-samples t-tests), I performed the non-parametric versions of the independent and the paired t-tests. These were the Mann Whitney U test (also called Wilcoxon rank-sum test) and the Wilcoxon signed-rank test, respectively. Both tests were fitted with the function *wilcox.test* from the *stats* package. The standardised measure of effect size computed was r, fitted with the function *wilcox_effsize* from the *rstatix* package. An r of 0.1-0.3, 0.3-0.5 and ≥ 0.5 was considered small, medium and large, respectively. The results of these tests will be reported in a footnote in the results section.

translations of the Galician exposure and test verbs. Responses were coded as binary (1 = the meaning of a verb was known, 0 = it was unknown). If participants reported not knowing the meaning of a verb, they were asked whether they had heard that verb before (1 = they had heard it, 0 = they had not) and whether they would be able to interpret it in the context of a sentence (1 = they would, 0 = they would not). I calculated the percentage of exposure and test verbs that participants in each experiment reported knowing the meaning of, having heard before and being able to interpret in a sentence.

2.7. Results

In this section, I present the results of Experiment 3 and part of the results of Experiment 2. The latter were already presented in Chapter 2, but are repeated below to ease the comparison with Experiment 3. Henceforth, the group of participants who processed sentences with high frequency verbs (Experiment 2) will be referred to as the *high frequency group* and the group of participants who processed sentences with low frequency verbs (Experiment 3) will be referred to as the *low frequency group.*

2.7.1. Testing phase

The similar and the dissimilar structure as part of the linguistic system

The mean accuracy percentages for the high frequency group and the low frequency group when judging the similar structure (SS), the dissimilar structure (DS) and the structures violating them (*SS and *DS) are presented in Table 3.3. As shown, the two groups judged significantly more than 50% of all conditions correctly. This shows that the two could distinguish the structures that were correct in the language from the structures that were not. Despite participants' accurate performance, Chapter 2 showed that the high frequency group was biased towards judging as "correct" both SS and *SS items, and DS and *DS items, regardless of whether the structures were grammatical or not. This was indicated by significantly below zero mean indices of response bias *c* (cf. SS and *SS items, M = -0.16, SD = 0.35; t (43) = -3.01, p < .01, d = -0.15; DS and *DS items, M = -0.10, SD = 0.32; t (43) = -2.01, p = .03, d = -0.30). The same indices were calculated for the low frequency group and a similar response bias was found (SS and *SS items, M = -0.15, SD = 0.24; t (43) = -4.08, p < .001, d = -0.62; DS and *DS items, M = -0.19, SD = 0.30; t (43) = -4.10, p < .001, d = -0.62). A *d′* analysis assessed sensitivity to the difference between grammatical conditions and their ungrammatical counterparts separate from response bias. As a reminder, a *d′* score of 0 reflects no discrimination between grammatical and ungrammatical sentences (0% sensitivity) and a *d′* score of 4.65 reflects excellent discrimination between the two (100% sensitivity) (Macmillan & Creelman, 2005). The high frequency group and the low frequency group were sensitive to the difference between SS and *SS conditions and between DS and *DS conditions. Specifically, the mean *d′* scores assessing discriminability between the two pairs of structures were significantly above zero for the two groups of participants (Table 3.4).

		SS	*SS	DS	*DS
High freq. group	*M*	89.26***	78.25***	82.29***	76.57***
	SD	30.98	41.28	38.20	42.38
	95%CI	[87.20, 91.31]	[75.51, 80.98]	[79.75, 84.82]	[73.76, 79.38]
Low freq. group	*M*	83.01***	71.79***	73.86***	60.02***
	SD	37.58	45.03	43.96	49.01
	95%CI	[80.52, 85.50]	[68.81, 74.77]	[70.96, 76.77]	[56.78, 63.27]

TABLE 3.3. Mean accuracy (%), standard deviations (%) and 95% confidence intervals for all conditions in the GJT conducted by the high frequency group (Experiment 2) and the low frequency group (Experiment 3). *M* = Mean, *SD* = Standard Deviation, *95%CI* = 95% Confidence Interval. Significance from chance: *** $p < .001$.

		SS vs. *SS	DS vs. *DS
High freq. group	*M*	2.42***	2.06***
	SD	1.52	1.58
	95%CI	[2.35, 2.49]	[1.99, 2.14]
Low freq. group	*M*	1.98***	1.25***
	SD	1.68	1.77
	95%CI	[1.91, 2.06]	[1.72, 1.34]

TABLE 3.4. Mean *d'* scores, standard deviations and 95% confidence intervals reflecting sensitivity to SS vs.*SS and DS vs.*DS in the GJT conducted by the high frequency group (Experiment 2) and the low frequency group (Experiment 3). *M* = Mean, *SD* = Standard Deviation, *95%CI* = 95% Confidence Interval. Significance from zero: ***$p < .001$.

In sum, the accuracy and the *d'* analyses show that the high frequency group and the low frequency group could recognize that the similar and the dissimilar structure were correct in the L2 and that the structures violating them were not. This evidences that the two grammatical structures were present in both groups of learners' linguistic systems.

Comparing the establishment of the similar and the dissimilar structure in the linguistic system

Figure 3.2 represents mean accuracy percentages for the high frequency group and the low frequency group organized by structure: on the one hand, accuracy when judging the similar structure and its ungrammatical counterpart; on the other hand, accuracy when judging the dissimilar structure and its ungrammatical counterpart. Figure 3.3 represents mean accuracy percentages when judging the two pairs of structures organized by group of participants: on the one hand, accuracy for the high frequency group and, on the other hand, accuracy for the low frequency group. As advanced in Section 2.6.1, a generalized linear mixed effects model tested for the interaction between the effect of Cross-linguistic similarity (Similar structure vs.

Dissimilar structure) and Verb frequency (High vs. Low) on accuracy. This yielded an effect of Cross-linguistic similarity (β = 0.56, *SE* = 0.07, *z* = 7.46; *p* < .001), no effect of Verb frequency (β = 0.72, *SE* = 0.45, *z* = 1.60; *p* = .11) and an interaction between the two variables (β = -0.33, *SE* = 0.15, *z* = -2.19; *p* = .029). In light of this interaction, comparisons by verb frequency and cross-linguistic similarity were conducted.

On the one hand, as illustrated in Figure 3.2, the high frequency group correctly judged a mean of 83.74% (*SD* = 36.91%, 95%CI = [82.01, 85.47]) of all SS and *SS items, while mean accuracy for the low frequency group was somewhat lower, 77.39% (*SD* = 41.84%, 95%CI = [75.43, 79.35]). In spite of this, as predicted by H1, accuracy when judging SS and *SS items was statistically comparable for the two groups (β = 0.50, *SE* = 0.54, *z* = 0.92; *p* = .36). The estimated coefficient of the effect of Verb frequency on accuracy (in log odds) was 0.50, which corresponds to an odds ratio of 1.65 to 1. That is, the odds of judging a SS or *SS item correctly as opposed to incorrectly were 1.65 times higher when a participant was in the high frequency group compared to the low frequency group, but this effect was not significant[14]. Turning to DS and *DS sentences, learners in the high frequency group correctly judged a mean of 79.43% (*SD* = 40.43%, 95%CI = [77.53, 81.32]) of these items, while mean accuracy for learners in the low frequency group was lower, 66.95% (*SD* = 47.05%, 95%CI = [64.75, 69.15]). This time, as predicted by H2, accuracy when judging DS and *DS items was significantly higher for the high frequency group than for the low frequency group (β = 0.93, *SE* = 0.42, *z* = 2.23; *p* = .026). The estimated logit coefficient of the effect of Verb frequency on accuracy was 0.93 (odds ratio: 2.53 to 1). In other words, the odds of judging a DS or *DS item correctly as opposed to incorrectly were 2.53 times higher when a participant was in the high frequency group compared to the low frequency group and this effect was significant.

[14] I calculated odds ratio using the function *exp()* from the R *base* package (R Core Team, 2022).

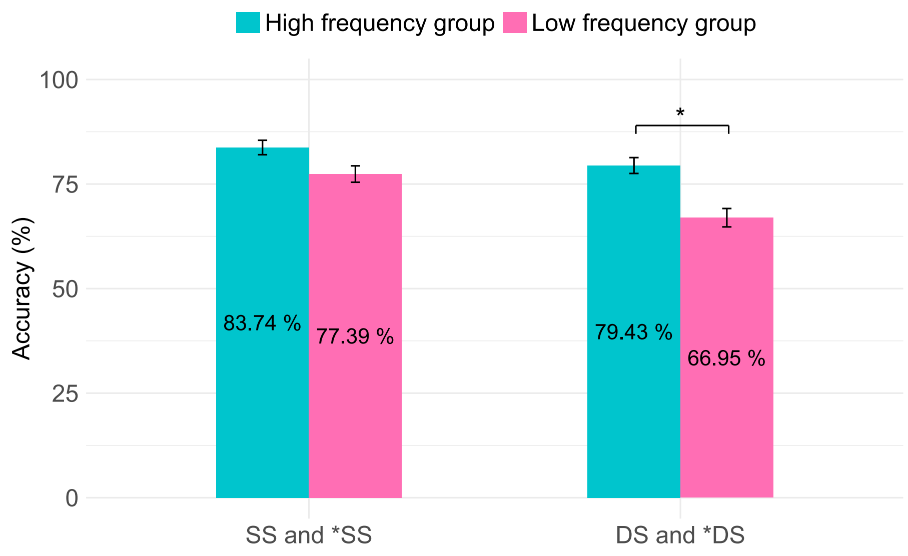


FIGURE 3.2. Mean accuracy (%) for the high frequency group (Experiment 2) vs. the low frequency group (Experiment 3) in SS and *SS conditions and DS and *DS conditions. Error bars represent 95% confidence intervals.

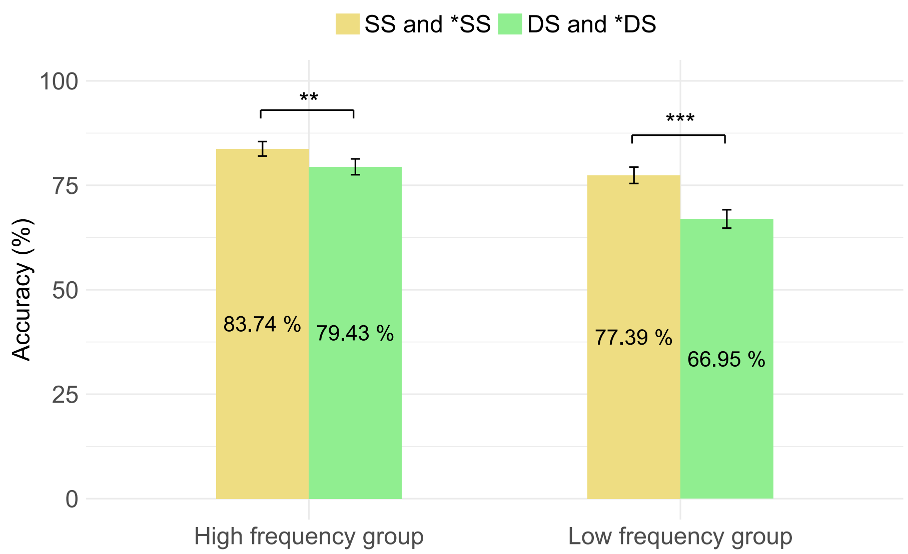


FIGURE 3.3. Mean accuracy (%) in SS and *SS conditions vs. DS and *DS conditions for the high frequency group (Experiment 2) and the low frequency group (Experiment 3). Error bars represent 95% confidence intervals.

On the other hand, as shown in Figure 3.3, the high frequency group was significantly more accurate when judging SS and *SS items than DS and *DS items ($\beta = 0.61$, $SE = 0.20$, $z = 3.08$; $p < .01$). This was also the case for the low frequency group ($\beta = 1.27$, $SE = 0.16$, $z = 7.87$; $p < .001$). Importantly, in line with the predictions of H3, the magnitude and strength of the difference between the two pairs of structures was smaller for the high frequency group than for the low frequency group. In the first group, the estimated logit coefficient of the effect of Cross-linguistic similarity on accuracy was 0.61, which corresponds to an odds ratio of 1.84 to 1. In brief, the odds that participants in the high frequency group judged a sentence correctly as opposed to incorrectly were 1.84 times higher when the item was SS or *SS compared to when it was DS or *DS. Conversely, in the second group, the estimated logit coefficient of the effect of Cross-linguistic similarity on accuracy was 1.27 (odds ratio: 3.56 to 1). That is, the odds that participants in the low frequency group judged a sentence correctly as opposed to incorrectly were 3.56 times higher when the item was SS or *SS compared to when it was DS or *DS. In addition, the two effects were significant but, for the high frequency group, the p-value was smaller than .01 and, for the low frequency group, it was smaller than .001. This means that the evidence in favour of a true difference between the two pairs of structures (SS and *SS vs. DS and *DS) was stronger for the low frequency group than for the high frequency group (Winter, 2020).

The *d′* analysis corroborated the results of the accuracy analysis. Figure 3.4 shows the boxplots illustrating the distribution of the *d′* scores for the high frequency group and the low frequency group organized by structure: on the one hand, scores reflecting sensitivity to the difference between the similar structure and its ungrammatical counterpart; on the other hand, scores reflecting sensitivity to the difference between the dissimilar structure and its ungrammatical counterpart. Figure 3.5 represents the boxplots of the distribution of the *d′* scores reflecting sensitivity to the difference between the two pairs of structures organized by group of participants: for the high frequency group and for the low frequency group. A 2x2 ANOVA analysed whether there was an interaction between Cross-linguistic similarity and Verb frequency on *d′* scores. The test indicated that there was a main effect of Cross-linguistic similarity (F (1, 86) = 50.72, p < .001, large effect size of η_p^2 = .371), no main effect of Verb frequency (F (1, 86) = 3.25, p = .08, η_p^2 = .036) and an interaction between the two factors (F (1, 86) = 6.05, p = .02, medium effect size of η_p^2 = .066). Because of this interaction, comparisons by verb frequency and cross-linguistic similarity were carried out. On the one hand, independent-samples t-tests indicated that the mean *d′* score for SS vs. *SS did not statistically differ between the high frequency group and the low frequency group (t (86) = 1.26, p = .21, d = 0.27). That is, as predicted by H1, the two groups of participants were comparably sensitive to the difference between the similar structure and its ungrammatical counterpart. By contrast, the mean *d′* score for DS vs. *DS was significantly larger for the high frequency group than for the low frequency group (t (86) = 2.24, p = .028, small effect size of d = 0.48). As predicted by H2, the high frequency group was significantly more sensitive to the difference between the dissimilar structure and its ungrammatical counterpart than the low frequency group.

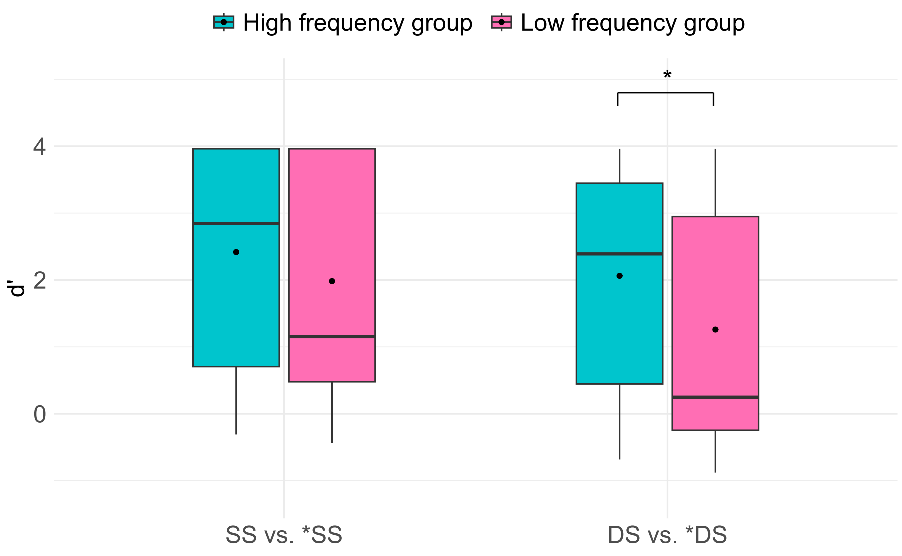


FIGURE 3.4. Distribution of *d′* scores for the high frequency group (Experiment 2) compared to the low frequency group (Experiment 3) in SS vs. *SS conditions and DS vs. *DS conditions. The black horizontal line in the box of the boxplot shows the median. The black dot is the mean. Whiskers end at the smallest and largest data points that fall within 1.5 times the interquartile range from the first quartile (25%) and the third quartile (75%) of the data.

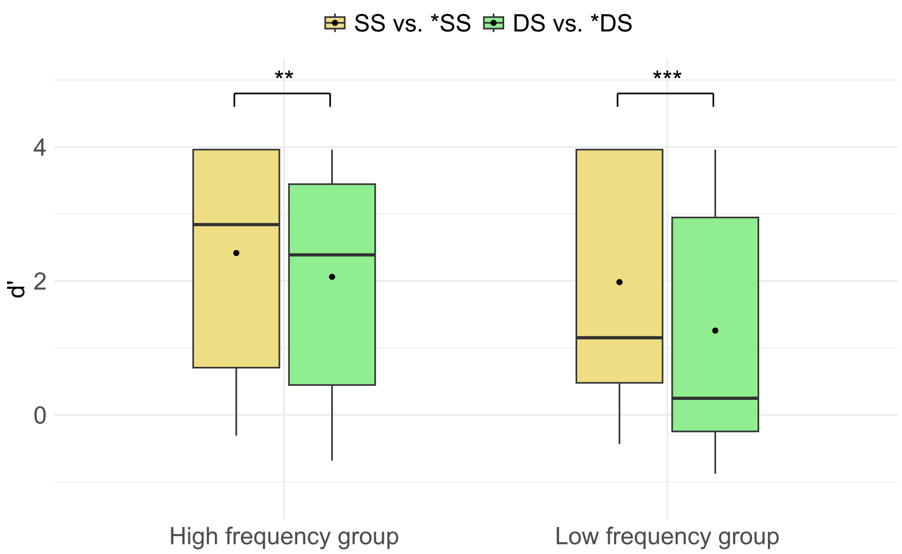


FIGURE 3.5. Distribution of *d′* scores in SS vs. *SS conditions compared to DS vs. *DS conditions for the high frequency group (Experiment 2) and the low frequency group (Experiment 3). The black horizontal line in the box of the boxplot shows the median. The black dot is the mean. Whiskers end at the smallest and largest data points that fall within 1.5 times the interquartile range from the first quartile (25%) and the third quartile (75%) of the data.

In addition, paired-samples t-tests revealed that the mean *d′* score indicating sensitivity to SS vs. *SS was significantly larger than the mean *d′* score indicating sensitivity to DS vs. *DS, both for the high frequency group (t (43) = 3.06, p = .004, medium effect size of d = 0.50) and for the low frequency group (t (43) = 7.40, p <.001, large effect size of d = 1.12)[15]. As described in the accuracy analysis, even if sensitivity to SS vs. *SS was larger than to DS vs. *DS for both groups of participants, the magnitude and strength of the difference was larger for the low frequency group than for the high frequency group, as predicted by H3. On the one hand, the standardized measure of effect size (Cohen's d) indicated that the size of the effect was larger for participants in the low frequency group (d = 1.12) than the high frequency group (d = 0.50). On the other hand, the p-value indicating that there was a true difference between sensitivity to the two pairs of structures was below the threshold of .01 for the high frequency group and below the threshold of .001 for the low frequency group. This indicates that the evidence in favour of this difference was stronger for the low frequency group than for the high frequency group.

In sum, the accuracy and *d′* analyses suggest that, in line with H1, the similar structure was comparably established in the linguistic system as a result of processing sentences with high frequency verbs and low frequency verbs. Conversely, in line with H2, the dissimilar structure was more robustly established in the linguistic system as a result of processing sentences with high frequency verbs than low frequency verbs. Finally, following H3, the similar structure was more robustly established in the linguistic system than the dissimilar structure irrespective of verb frequency, but the advantage of the first structure over the second was reduced when processing sentences with high frequency verbs.

2.7.2. Debriefing phase

2.7.2.1. Verbal report

As shown in Chapter 2 (Section 3.7.2), only those participants in the high frequency group who learnt the similar and the dissimilar structure could potentially report them. The same occurred in Experiment 3 for the low frequency group. In this light, I report awareness of the similar and the dissimilar structure for those participants in the high frequency group and the low frequency group whose performance in the GJT evidenced learning of one or the two structures. These participants judged SS and *SS items and/or DS and *DS items at or above 65% accuracy (clearly above chance) and their *d′* scores indicating sensitivity to the difference between SS and *SS and/or between DS and *DS were greater than zero. In the high frequency group, 73% (32/44) of participants learnt one or the two structures. Out of these, 88% (28/32)

[15] Non-parametric tests yielded the same results as their parametric counterparts:
High frequency group compared to low frequency group (SS vs. *SS): W = 1069.5, p = .39, small effect size of r = 0.09.
High frequency group compared to low frequency group (DS vs. *DS): W = 1256, p = .016, small effect size of r = 0.26.
SS vs. *SS compared to DS vs. *DS (High frequency group): V = 614.5, p <.01, medium effect size of r = 0.49.
SS vs. *SS compared to DS vs. *DS (Low frequency group): V = 695, p <.001, large effect size of r = 0.82.

could provide metalinguistic descriptions of the structures and, thus, were considered *aware* of them. Specifically, all aware participants reported the similar and the dissimilar structure, except P19 and P32 who, in spite of learning the two structures, verbalized just the similar one[16]. The remaining four participants were considered *unaware* of the structures*:* one (P31) did not answer to the verbal report and the other three just reported the presence/absence of *que* in the structures (P8) or the presence of a finite or non-finite embedded verb (P33 and P38). In the low frequency group, 61% (27/44) of participants learnt one or both structures according to their performance in the testing phase and 78% (21/27) of these were coded as *aware.* All aware participants learnt and reported the similar and the dissimilar structure, except one participant (P2), who only learnt and verbalized the similar structure. The remaining six participants were coded as *unaware.* Specifically, two (P5 and P38) identified that the varying elements in the structures were the presence/absence of *que* and a finite or non-finite verb, but could not correlate them. Three participants (P20, P35 and P44) simply reported that some sentences contained *que* and others did not. Finally, P32 provided an answer unrelated to the target structures, i.e. reported noticing that in some sentences the determiner before the inanimate noun was *a* ("the", feminine) and in others it was *o* ("the", masculine).

2.7.2.2. Vocabulary test

Of all participants in the high frequency group, 68% (30/44) agreed to carry out the online version of the vocabulary test. They all reported knowing the meaning of the Spanish translations of all exposure and test verbs. As for participants in the low frequency group, they reported knowing the meaning of a mean of 98.07% (*SD* = 13.77%, 95%CI = [97.16, 98.98]) of the Spanish translations of exposure verbs and 96.02% (*SD* = 19.55%, 95%CI = [94.73, 97.32]) of the Spanish translations of test verbs. Those participants who reported not knowing the meaning of some verbs reported having heard them all before and feeling like they could interpret them in the context of a sentence. Overall, I take this to be good-enough evidence that participants in Experiments 2 and 3 were familiar with the Spanish counterparts of the Galician cognate verbs used in these experiments.

2.8. Discussion

Chapter 3 investigated, for the first time, whether differences in lexical frequency, as in the case of high frequency vs. low frequency words, facilitate the initial acquisition of cross-linguistically similar and dissimilar L2 structures. Specifically, in this chapter I compared how Spanish natives without knowledge of Galician learnt the *similar structure*, existing in Spanish and Galician, and the *dissimilar structure*, conveying the same meaning but existing only in Galician, following exposure to these structures in sentences with high frequency verbs (Experiment 2, high frequency group) and low frequency verbs (Experiment 3, low frequency group). Like in Experiment 2, in Experiment 3 participants were exposed to the structures via

[16] P = Participant.

a structure-search task and, then, they were tested on their learning of the structures via a Grammaticality Judgement Task (GJT) with feedback. All sentences in the exposure and testing phases were made up of Spanish-Galician cognates. Learning of the structures could occur both in the exposure phase and in the testing phase, due to feedback. Thus, I manipulated the lexical frequency of the Spanish translations of the Galician verbs in the exposure and testing phases, so that it was lower than in Experiment 2. I assumed that since participants had never been exposed to Galician, in Experiment 2 and in Experiment 3 the L2 verbs would be processed as if they were equivalent to the high or low frequency L1 verbs. A vocabulary test evidenced that participants in both experiments were familiar with the Spanish counterparts of the Galician verbs. Finally, like Experiment 2, Experiment 3 included a verbal report. I postulated three hypotheses regarding how differences in lexical frequency would affect acquisition of L2 structures in Experiment 2 vs. 3. These were based on how MOGUL suggests that high frequency and low frequency words are stored and processed, how acquisition by processing of cross-linguistically similar and dissimilar L2 structures may take place and how lexical and syntactic processing interact during this process.

On the one hand, I hypothesized that the acquisition of the dissimilar structure would be facilitated by the stronger activation of high frequency verbs (Experiment 2) compared to low frequency verbs (Experiment 3). Specifically, I argued that the higher the activation of a verb was during processing, the higher would be the activation of the structure containing it and the greater would be the learning of the structure resulting from processing. As evidence in favour of this hypothesis, accuracy when judging the grammatical and the ungrammatical dissimilar structure in the GJT was significantly higher for the high frequency group than for the low frequency group. The same occurred for the *d′* scores indicating sensitivity to the difference between the two structures. On the other hand, I hypothesized that the acquisition of the similar structure would not be facilitated by differences in lexical frequency; this structure would be processed using an L1 structure firmly established in the linguistic system and, thus, would be non-significantly affected by verb frequency. This hypothesis was also confirmed. In the GJT, accuracy when judging the grammatical and the ungrammatical similar structure was comparable for the high frequency group and the low frequency group, and so was sensitivity to the difference between the two structures. Finally, I hypothesized that, like in Experiment 2, in Experiment 3 there would be a learning advantage for the similar structure over the dissimilar one. Nevertheless, I expected this advantage to be smaller when learning the structures with high frequency verbs than with low frequency verbs. In line with this, in the GJT overall accuracy for the similar structure and its ungrammatical counterpart was significantly higher than for the dissimilar structure and its ungrammatical counterpart, and so was sensitivity to the difference between the first two structures compared to the second. Yet, the magnitude and the strength of this advantage was smaller for the high frequency group than for the low frequency group. In what follows, I discuss how the two groups could have processed the structures during the exposure and the testing phases, and how this could have yielded the results observed. Finally, I discuss the results of the debriefing phase.

2.8.1. Discussion of the exposure and testing phases

I propose that sentences were processed in the same way in the exposure and testing phases. Based on MOGUL (Sharwood Smith & Truscott, 2014, sec. 4.2, 10.3), I assume that when learners in the high frequency group and the low frequency group encountered a sentence formed by the similar structure, the equivalent L1 structure activated and was used to process the input. On the other hand, when learners first encountered a sentence formed by the dissimilar structure in the exposure phase, an appropriate syntactic representation was created, simply as a means of processing the input. This received a low resting activation level. As this structure was subsequently processed as part of the L2 in the exposure and testing phases, its resting activation level increased and it gradually became established in learners' linguistic system. As discussed in Chapter 2, the view that the syntactic representations of the L1 are present at the beginning of L2 acquisition and, whenever possible, are used to process the L2, but that L2 syntactic representations not shared with the L1 have to be acquired from input, is compatible with several theories and models of L2 acquisition. These include the Full Transfer/Full Access model (B. D. Schwartz & Sprouse, 1994, 1996), the Autonomous Induction Theory (Carroll, 1999), the Unified Competition Model (MacWhinney, 2005) and the Micro-cue model of L2 acquisition (Westergaard, 2021).

Importantly, I argue that when Spanish natives with no knowledge of Galician read and listened to the Galician cognate embedded verbs (and, more generally, to all cognate words), the orthographic and phonological representation of their Spanish translations activated. These activated coindexed syntactic and conceptual representations, which, in turn, were coindexed with the orthographic and phonological representation of the Galician verbs in order to process them (see this chapter, Section 1.3.4). The assumption that when processing cognate words their counterparts in the non-target language activate is taken to be responsible for the finding that cognates are processed faster and more accurately than non-cognates (e.g. Dijkstra et al., 1999; Dijkstra & van Heuven, 1998; Lemhöfer & Dijkstra, 2004; van Hell & Dijkstra, 2002, see Chapter 1, Section 2 for more details). Based on MOGUL, I assume that the Spanish high frequency verbs that activated upon reading and listening to the Galician cognate verbs had a higher resting activation level than the Spanish low frequency verbs. Consequently, the current activation level of the former, defined as the sum of the resting activation level and activation received during the processing event, was higher than that of the latter (Sharwood Smith & Truscott, 2014, sec. 3.3.5, see also Figure 3.1). This current activation spread from the chain of representations of the L1 high or low frequency verbs to coindexed representations of the L2 verbs, causing the Galician translations of Spanish high frequency verbs to have a higher current activation level than the Galician translations of Spanish low frequency verbs. The idea that the frequency of words in the L1 is available when processing words with similar form in the L2 resonates with the finding that, during interlingual homograph recognition in the L2, the higher frequency of the L1 reading of the homograph compared to its L2 reading slows down recognition time (Dijkstra et al., 1998).

I propose that when learners processed the embedded verb in sentences formed by the similar or the dissimilar structure, the verb's syntactic representation activated. This activation spread to the representation of the structure including it, which caused its current activation level to rise. The rise in the current activation level of the similar or the dissimilar structure depended on the current activation level of the verb (Sharwood Smith & Truscott, 2014, sec. 3.4.1). Hence, the rise was larger when activation spread from a Galician verb cognate with a Spanish high frequency verb than from a Galician verb cognate with a Spanish low frequency verb. When sentence processing terminated, the current activation level of the structures fell back towards its resting activation level, landing at a position slightly above the original. In other words, the resting activation level of the structures increased each time they were processed. Representations with a high current activation level land at a higher resting activation level than representations with a lower current activation level (Sharwood Smith & Truscott, 2014, sec. 3.3.5). Consequently, the similar and the dissimilar structure should have had a higher resting activation level when processed with high frequency verbs than with low frequency verbs.

Crucially, I assume that the similar structure, shared for the L1 and the L2, was stored in learners' linguistic system with a high resting activation level, resulting from it being previously processed in the L1. By contrast, the dissimilar structure, which learners had never processed before, was initially attributed a low resting activation level. Each time that the similar and the dissimilar structure were processed, their resting activation level increased. Yet, MOGUL assumes that the increases in resting activation level as a function of processing are logarithmic. A representation's resting activation level increases quickly when it is low but, as the resting level rises, the increases derived from processing diminish until, at some point, the resting level is so high that it does not increase any more (Sharwood Smith & Truscott, 2014, sec. 4.6.5). Because I assume that the resting activation level of the similar structure was high from the moment L2 learners first encountered it, I hypothesize that the increases resulting from additional processing were small. Thus, even if the resting activation level of the similar structure increased marginally more for the high frequency group than for the low frequency group, I assume that, overall, the resting level of the structure was comparable between groups. Simply put, I argue that the similar structure was comparably established in learners' linguistic system irrespective of the frequency of the verbs it was processed with. By contrast, because the resting activation level of the dissimilar structure was initially low, the increases derived from additional processing should have been quite large. Consequently, I assume that the larger increases in the structure's resting activation level resulting from processing sentences with high frequency verbs than low frequency verbs caused the dissimilar structure to be more firmly established in the linguistic system of the high frequency group than the low frequency group. This would explain the results of the GJTs in Experiment 2 vs. 3.

I hypothesize that, because the (grammatical) similar and dissimilar structures were established in learners' linguistic system, they generally judged them as correct in the L2. As for the ungrammatical similar and dissimilar structures, I hypothesize that learners processed them as if they were their grammatical counterparts, incrementally activating the representation of these structures, until processing the embedded verb, where the violation took place. Since the similar and the dissimilar structure were firmly established in learners' linguistic system, this facilitated identifying the mismatch between these structures and their ungrammatical counterparts. The fact that the dissimilar structure was more robustly established in the linguistic system of the high frequency group than the low frequency group would explain why accuracy for the grammatical and the ungrammatical dissimilar structure and sensitivity to the difference between the two were significantly higher for the first group than for the second. Likewise, the fact that the resting activation level of the similar structure increased slightly more when processed with high frequency verbs than with low frequency verbs could explain why, descriptively, accuracy for the grammatical and the ungrammatical similar structure and sensitivity to the difference between the two was somewhat higher for the high frequency group than for the low frequency group. However, this accuracy and sensitivity did not statistically differ between groups. This goes in line with the hypothesis that the similar structure was comparably established in the linguistic system of learners in the high and the low frequency groups. Future studies could investigate whether, if a different task or method were used to assess learning, a small but significant facilitation would be observed when learning the similar structure with high frequency verbs compared to low frequency verbs.

Finally, as mentioned, I assume that the similar structure had a higher resting activation level than the dissimilar structure due to its prior processing in the L1, irrespective of whether these structures were processed with high frequency or low frequency verbs. Additionally, I assume that the similar structure was comparably established in the linguistic system when processed with high frequency verbs and with low frequency verbs, but that the dissimilar structure was more firmly established in the linguistic system when processed with high frequency verbs. That being so, I argue that the advantage of the similar structure over the dissimilar structure was smaller when processing the structures with high frequency verbs than with low frequency verbs. This would explain why learners in both the high frequency group and the low frequency group were significantly more accurate when judging the grammatical and the ungrammatical similar structure than the grammatical and the ungrammatical dissimilar structure, and why they were significantly more sensitive to the difference between the first pair of structures than the second. Importantly, it would also explain why this difference in accuracy and sensitivity was less pronounced for the high frequency group than for the low frequency group.

2.8.2. Discussion of the debriefing phase

2.8.2.1. Verbal report

Like Experiment 2, Experiment 3 used an explicit learning paradigm, which encouraged learners to think about the target syntactic structures. On the one hand, at the beginning of the exposure phase learners were informed that they would be exposed to sentences formed by two different structures and they were instructed to try to discover what these structures were. On the other hand, the testing phase was a GJT with feedback. This could have caused learners to consciously focus on the form of the sentences to find out, if necessary, why feedback indicated some to be correct in the L2 and others not (Leeman, 2007). Explicit learning paradigms tend to produce explicit (i.e. verbalizable) knowledge of the structures (N. C. Ellis, 1993; Rebuschat, 2009; Robinson, 1997; Tagarelli et al., 2016). Accordingly, almost 90% of learners in the high frequency group who learnt the structures could verbalize them and, thus, were *aware* of them. The percentage was somewhat lower for the low frequency group (almost 80%). Nevertheless, both percentages were well above 50%, which indicates that the majority of learners in the high frequency group and the low frequency group who learnt the structures had explicit or conscious knowledge of them. Two participants in the high frequency group and one in the low frequency group verbalized only the similar structure. On the one hand, there was evidence that the two participants in the high frequency group had learnt both the similar and the dissimilar structure. As discussed in Chapter 2 (Section 3.8.3), these participants did not report the dissimilar structure due to a technical error when conducting the verbal report. On the other hand, for the participant in the low frequency group, there was only evidence that s/he had learnt the similar structure, so it is possible that s/he did not report the dissimilar structure simply because s/he had not learnt it.

Four learners in the high frequency group and six learners in the low frequency group learnt the similar and/or the dissimilar structure but could not report them and, thus, were considered *unaware* of them. Some of these unaware learners reported that sentences varied in the presence/absence of *que* or in the form of the embedded verb. As discussed in Chapter 2, it could be that these participants did not provide an accurate description of the similar and the dissimilar structure because they truly had no explicit knowledge of them or it could be that they had some conscious knowledge of the structures, but not the ability to describe them with words. An unaware participant in the high frequency group did not respond to the verbal report. It could be that this learner had conscious knowledge of the structures but was not confident enough to report it, or simply decided not to answer. Finally, a participant in the low frequency group did not provide a description of the similar and the dissimilar structure, in spite of having learnt them, but instead verbalized the two possible determiners in the experimental sentences. It could be that the paradigm resulted in unconscious knowledge of the structures for this learner. This is not entirely unexpected since, as mentioned in Chapter 1 (Section 1.3), explicit learning paradigms sometimes result in explicit knowledge for some learners and in implicit knowledge for other learners (e.g. Rebuschat, 2009, Experiment 6; Robinson, 1997).

2.8.2.2. Vocabulary test

The thirty participants in the high frequency group who took the online version of the vocabulary test indicated that they knew the meaning of the Spanish translations of all exposure and test verbs (i.e. they were familiar with them). I assume that the remaining participants would have reported the same. Similarly, participants in the low frequency group indicated that they knew the meaning of the Spanish translations of over 95% of exposure and test verbs. They could have reported not knowing the meaning of the rest of verbs because they were unsure about what this meaning was or because they did not have a clear, concrete semantic representation of them. Following Hopp (2016), who conducted a similar vocabulary test for the high and low frequency verbs used in his experiment, I do not consider this a problem for the current study, because its aim was not to test comprehension of the target meaning of verbs or sentences. More importantly, for those verbs participants reported not knowing the meaning of, they indicated that they had heard them before and that they would be able to interpret them if they appeared in the context of a sentence. This suggests that participants accessed a syntactic and a semantic representation of these verbs when encountering their Galician cognate counterparts in the experiment, which was crucial for the frequency manipulation proposed to be effective.

3. Concluding remarks

By comparing novice adult L2 learners' acquisition of cross-linguistically similar and dissimilar structures processed in sentences with high frequency verbs (Experiment 2) and low frequency verbs (Experiment 3), I present evidence that helps understand, for the first time, how lexical frequency interacts with syntactic processing during initial L2 syntax acquisition. Specifically, I propose that the stronger activation of high frequency verbs compared to low frequency verbs facilitated the acquisition of a cross-linguistically dissimilar structure, which needed to be incorporated into the linguistic system from input. By contrast, my results indicate that the facilitative effect of lexical frequency did not obtain for the structure existing in the L1 and the L2, which according to several approaches to L2 acquisition would have been established in the linguistic system by the time L2 acquisition started and, strictly speaking, did not need to be acquired. A consequence of this is that, while the learning advantage for the cross-linguistically similar structure over the cross-linguistically dissimilar one observed in Chapter 2 replicated in Chapter 3, it was larger when the structures were processed with low frequency verbs. In other words, I argue that the stronger activation of high frequency verbs compared to low frequency verbs reduced the learning distance between the similar structure and the dissimilar one.

To my knowledge, no model, theory or framework of L2 processing and/or acquisition explicitly explains these findings. I propose that my results are predicted and can be interpreted within the MOGUL framework, based on how it suggests that high frequency and low frequency words are stored and processed, how structures differing in cross-linguistic similarity are processed and/or acquired and how lexical and syntactic processing interact in real time. Due to the exploratory nature of this work, further research needs to be conducted to replicate these findings and, more generally, to provide more evidence for the effect of word activation on the acquisition of cross-linguistically similar and dissimilar L2 structures. A possibility would be to manipulate the degree of lexical activation by means of a variable other than frequency, such as cognateness. This is addressed in Chapter 4.

Chapter 4

The facilitative role of cognates in initial L2 syntax acquisition*

1. Introduction

Bilinguals process cognate words faster and more accurately than non-cognate words, both in the L1 and the L2 (the *cognate facilitation effect,* CFE). Over the years, several explanations for this effect have been proposed. As mentioned in Chapter 1, perhaps the most cited and comprehensive one conceives this facilitation as the result of greater activation for cognates than non-cognates. More precisely, it has been argued that only cognates activate shared orthography and/or phonology across languages and that that activation spreads to a shared meaning representation, which then feeds back its activation to the word's orthographic and phonological forms. Consequently, the orthographic, phonological and semantic representations of cognates are more activated than those of non-cognates (Dijkstra & van Heuven, 2002). Chapter 3 presented evidence which suggests that the stronger activation of high frequency words compared to low frequency words might facilitate initial L2 syntax acquisition, but only when L2 structures are not similar to L1 grammar. Chapter 4 further explores whether differences in the activation of lexical items facilitate initial L2 syntax learning. Specifically, it investigates whether the stronger activation of cognates compared to non-cognates eases the acquisition of L2 structures and whether this facilitation is modulated by cross-linguistic syntactic similarity, as in Chapter 3.

Previous research indicates that cognates are recognized, read and produced faster and/or more accurately than non-cognates in several word processing tasks in the L1 and the L2, such as lexical decision tasks (e.g. L1, Fricke, 2022; van Hell & Dijkstra, 2002; L2, Dijkstra et al., 1999; Lemhöfer & Dijkstra, 2004, reading tasks using eye-tracking (e.g. L1, Cop et al., 2017; Van Assche et al., 2009; L2, Duyck et al., 2007; Van Assche et al., 2011) and picture-naming tasks (e.g. L1 and L2, Costa et al., 2000; Gollan et al., 2007), amongst others. In addition, word recognition and processing studies conducted with ERPs have shown that cognates elicit

* The experiments presented in this chapter were carried out in collaboration with Dr. Ruth de Diego-Balaguer during two research stays at the Department of Cognition, Development and Educational Psychology of the University of Barcelona (September-November 2021 and March-May 2022).

smaller amplitudes than non-cognates in the N400 time-window, associated with lexico-semantic processes such as word form-to-meaning mapping (e.g. Midgley et al., 2011; Peeters et al., 2013; Xiong et al., 2020). This provides additional evidence that cognates are easier to process than non-cognates.

To historical and contact linguists, cognates are words in two languages which have a similar or identical form and meaning and which share a common etymology, either because they derive from the same parent word (Schmitt, 1997) or because they descend from a common ancestor language (Whitley, 2002). In fact, the term *cognate* comes from the Latin word *cognatus*, which means "blood relative" (Otwinowska, 2016). Considering the etymological origin, cognates can be found, for instance, in typologically close Romance languages such as Spanish and Galician, which descended from Latin (cf. cognates in Experiments 1-3). Cognates may also be found in languages which are typologically more distant, such as Polish (Slavic language) and English (Germanic language), both Indo-European (e.g. Polish *mlyn* and English *mill*). However, these might share a lower degree of formal similarity and might be harder to recognize as cognates (Otwinowska, 2016). In addition, if cognates are defined considering just formal criteria —i.e. as words sharing form and meaning across languages— they might be found in unrelated languages, usually as loanwords (Ringbom, 2007). Examples of such loanwords are cognates in Spanish and Basque, which I study in this chapter. Basque has been in contact with other languages for more than 2000 years (first with Latin, later with Romance languages such as Spanish or French). Throughout these centuries, Basque has loaned a large number of features from these languages. For instance, it borrowed the participial suffix *–tu* from Latin. Then, it borrowed participial verbs from Spanish (e.g. *funcionado,* "function, work") and adapted them by replacing the regular ending *–do* by the suffix of perfective participle *–tu,* yielding cognate verbs such as *funtzionatu* ("function, work") (Hualde, 2000). In psycholinguistic studies, cognates are simply defined as words with similar orthography/phonology and meaning in two or more languages, irrespective of their etymological origin (e.g. Andras et al., 2022; Comesaña et al., 2018; de Groot & Comijs, 1995; Lemhöfer et al., 2008; van Hell & Dijkstra, 2002). This is also the definition adopted in this chapter. This requires specifying what *similar* orthography, phonology and meaning means.

Researchers have measured cross-linguistic formal similarity in several ways. For instance, Kroll and Stewart (1994) had native speakers of English without knowledge of German or Dutch translate a set of words from Dutch to English. If at least half of participants could translate a word, this was considered to have a sufficiently similar form in the two languages to be a cognate. In other studies (e.g. de Groot & Nas, 1991; Dijkstra et al., 1999, 2010; Tokowicz et al., 2002), bilinguals rated the formal similarity of pairs of translation equivalents on a 7-point scale (e.g. 1 = *no/low formal similarity*, 7 = *identical/high formal similarity*). More objective tools for assessing formal similarity are Van Orden's (1987) algorithm for orthographic similarity (e.g. Andras et al., 2022; Comesaña, Sánchez-Casas, et al., 2012; A. I.

Schwartz & Kroll, 2006; Van Assche et al., 2011)[1] and Levenshtein's distance for orthographic and phonological similarity (e.g. Andras et al., 2022; Comesaña et al., 2015; Cop et al., 2017; Schepens et al., 2012; *this dissertation*). As for cross-linguistic semantic similarity, it could be subjectively assessed in the same way as formal similarity, i.e. by asking bilinguals to rate translation equivalent pairs according to their similarity in meaning on a 7-point scale where 1 = *totally different* and 7 = *exactly the same* (e.g. Dijkstra et al., 1999, 2010; Tokowicz et al., 2002).

Some studies have evidenced that cognates might facilitate L2 syntax processing (e.g. X. Chen et al., 2023; Hopp, 2017; J. Huang et al., 2019; Soares et al., 2018, 2019; see Section 1.2 for more information). By contrast, as far as I know, whether and how differences in cognateness affect L2 syntax acquisition has not been investigated. To this aim, I conducted two experiments in which Spanish natives with no knowledge of Basque learnt a mini-language with Basque vocabulary and transitive structures designed to be either dissimilar (Experiment 4) or similar (Experiment 5) to Spanish. This was done by manipulating word order and marking of agent and patient arguments (verb-final word order and postpositional agent-patient marking in Experiment 4; verb-medial word order and prepositional patient marking in Experiment 5). In each experiment, two groups of participants learnt the structures by processing sentences with Spanish-Basque cognate or non-cognate verbs. As advanced in Chapter 1, to the best of my knowledge the existing models, theories and frameworks of L2 acquisition make no predictions about how embedding cognate or non-cognate words in cross-linguistically similar and/or dissimilar structures may influence their acquisition. I propose that the MOGUL framework may be appropriate to formulate hypotheses about this. In brief, following a similar reasoning to that in Chapter 3, I propose that the stronger activation of cognates compared to non-cognates should facilitate the acquisition of cross-linguistically dissimilar structures, which need to be learnt from input. By contrast, the facilitative role of cognates should be smaller, or even negligible, for cross-linguistically similar structures, which are already part of the linguistic system at the beginning of L2 acquisition. Experiments 4 and 5 provided evidence to retain these hypotheses.

Chapter 4 has the following organization. In Section 1.1, I review some studies supporting the cognate facilitation effect. Then, I detail how this effect is explained in the most well-known model of bilingual word processing, the BIA+ model (Dijkstra & van Heuven, 2002), and how it may be accounted for in MOGUL. In Section 1.2, I examine the influence of cognates on syntax processing, which is necessary to understand the influence that cognates might ultimately have on syntax acquisition. In Section 2, I focus on the effect of cognates on the acquisition of cross-linguistically dissimilar L2 structures. I present the structures to be learnt in Experiment 4 (Section 2.1) and I detail whether and how I hypothesize that cognates might influence the initial acquisition of these structures within MOGUL (Section 2.2). In Section 3, I

[1] Van Orden's algorithm calculates the orthographic similarity between two words as the ratio between the similarity of the two words' graphemes relative to the similarity of the target word's graphemes compared to itself.

describe and discuss Experiment 4. Next, in Section 4, I turn to the effect of cognates on the acquisition of cross-linguistically similar L2 structures. I present the structures to be learnt in Experiment 5 (Section 4.1) and I detail whether and how I propose that cognates might influence the initial acquisition of these structures within MOGUL (Section 4.2). In Section 5, I describe and discuss Experiment 5. The chapter ends with some concluding remarks in Section 6.

1.1. The influence of cognates on word processing

Evidence in favour of a processing advantage for cognates over non-cognates is found in written and oral word recognition, comprehension and production, both when words are presented in isolation and in context, in bilinguals' L1 and L2. The facilitation introduced by cognates has been accounted for differently, depending on the model of bilingual mental lexicon assumed and the representation of cognates presumed in that lexicon. Some proposals have explained the CFE in word recognition and comprehension with *identical cognates*, translation equivalents that are orthographically identical in two or more languages, e.g. *film-film* in English and Dutch (e.g. Kirsner et al., 1993; Lalor & Kirsner, 2000; Midgley et al., 2011; Peeters et al., 2013; Sánchez-Casas & García-Albea, 2005; Voga & Grainger, 2007) and *non-identical cognates*, translation equivalents with similar orthographic/phonological form in two or more languages, e.g. *tomato-tomaat* in English and Dutch (e.g. Dijkstra & van Heuven, 2002; Kirsner et al., 1993; Lalor & Kirsner, 2000; Sánchez-Casas et al., 1992; Sánchez-Casas & García-Albea, 2005). Other proposals have been advanced to account for the CFE in production (e.g. Costa et al., 2000; Kroll & Stewart, 1994; Sánchez-Casas & García-Albea, 2005; Strijkers et al., 2010). Since Chapter 4 investigates L2 syntax acquisition resulting from visual and spoken language processing, in this section I will focus on the CFE in visual and aural word recognition and processing, in and out of context. I will first review some experimental evidence in the L1 and the L2, to show that the effect is robust and bidirectional. Then, since the most detailed explanation of the CFE is the one proposed by the BIA+ model, I will discuss how the effect has been interpreted within this model. Given that in Experiments 4 and 5 I used non-identical cognates, I will focus on the CFE for this type of words. Finally, I will propose an account of the CFE within the MOGUL framework that is compatible with the explanation this effect receives in the BIA+ model.

The majority of studies examining the CFE in word recognition and processing have been conducted in the L2. One of the first and most influential studies was the one by Dijkstra et al. (1999). In a first experiment, Dutch-English bilinguals conducted a progressive demasking task. Stimuli were six types of English words, either cognates or false friends with Dutch depending on the similarity to their Dutch counterparts in terms of semantics, orthography and/or phonology. For each experimental word, an English non-cognate control word was selected. In the first trial of the experiment, participants saw a mask for 300ms, which was then replaced by a target word presented for 15ms. In subsequent trials, the presentation time of the mask gradually decreased, while that of the target word increased. Participants had to press a button when they identified the target word and then they had to write the word down in a text box. Reaction times and accuracy in word identification were calculated for each word type. Participants recognized English-Dutch cognate words overlapping in (i) semantics and orthography and (ii) semantics, orthography and phonology significantly faster and more accurately than non-cognate control words. In a second experiment, another group of Dutch-English bilinguals performed a visual lexical decision task. The same stimuli as in Experiment 1 were used, together with a set of nonwords constructed from the experimental English words. In each trial, participants saw a string of letters and had to indicate, as quickly as possible, whether that string was an English word or not. As in the previous experiment, decision time for cognates sharing semantics, orthography (and phonology) between English and Dutch was significantly shorter than for non-cognate controls, and responses were significantly more accurate for the former than for the latter. The authors concluded that cognates' cross-linguistic similarity in orthography and semantics facilitated word recognition, in line with the CFE.

A large number of studies have replicated this effect using similar paradigms. For instance, Lemhöfer and Dijkstra (2004) aimed to assess the reliability of the results obtained in Dijkstra et al.'s (1999) lexical decision task by conducting two separate lexical decision experiments, one testing just false friends (Experiment 1) and the other testing just cognates (Experiment 2). The participant profile, materials and procedure were the same as in the 1999 experiment. The results of the lexical decision task with cognates confirmed the validity of the data reported by Dijkstra and colleagues, for almost identical effects were obtained. That is, those cognates which had similar semantics and orthography and similar semantics, orthography and phonology were responded to significantly faster and more accurately than non-cognate control words[2]. Likewise, Lemhöfer (2008) used a progressive demasking task to study the impact of a number of lexical variables, amongst which cognate status, on word recognition. Participants were French, German and Dutch natives bilingual with English. The cognate words studied were orthographically identical between English and participants' L1. Non-cognates were translation equivalents with different spellings across languages. The procedure of the experiment was very similar to the one described for Dijkstra et al.'s (1999) Experiment 1. Participants saw a mask followed by an English word that needed to be identified and reported as soon as possible. The duration of the target gradually increased, while that of the mask shortened. The results of the study confirmed the CFE: the three groups of bilinguals identified cognates significantly faster than non-cognates. Similar results can be found in other experiments using progressive demasking tasks (e.g. Dijkstra et al., 2010) and, above all, visual

[2] In both Dijkstra et al. (1999) and Lemhöfer and Dijkstra (2004), the facilitative effect of cognates in word recognition time did not obtain for those translation equivalents sharing phonology (but not having identical orthography) in English and Dutch. This was attributed to a difference in the timing of activation of orthographic and phonological codes during word recognition. Orthographic information is available slightly earlier than phonological information. This arguably allowed participants to recognize the words on the basis of English orthographic information before phonological cross-linguistic competition could impact response times (see Lemhöfer and Dijkstra, 2004, for more information).

lexical decision tasks (e.g. Casaponsa et al., 2015; Dijkstra et al., 2010; Duyck et al., 2007; Krogh, 2022; Mulder et al., 2015; Peeters et al., 2013; Poort & Rodd, 2017; Vanlangendonck et al., 2020).

Research examining the CFE in the L1 is scarcer. Some experiments using visual lexical decision tasks suggest that cognates might also have an advantage over non-cognates in L1 visual word recognition. For instance, van Hell and Dijkstra (2002) conducted two experiments with trilingual speakers of L1 Dutch, L2 English and L3 French. In both experiments, stimuli were 20 Dutch-English cognates with similar orthography, phonology and meaning between languages, 20 Dutch-French cognates with the same characteristics and 40 Dutch items non-cognate with English or French. In addition, a set of pseudowords was created based on Dutch words. In the first experiment, participants had a higher proficiency in English than in French. Results showed that Dutch-English cognates were responded to significantly faster than non-cognates, but this was not the case for Dutch-French cognates. To test whether a stronger knowledge of the L3 was required to influence L1 processing, in the second experiment a group of trilinguals with a higher proficiency in French was selected. This time, significantly shorter response times were obtained for Dutch-English cognates and for Dutch-French cognates compared to non-cognates. Yet, other studies testing for the CFE in the L1 using visual lexical decision tasks have failed to find this effect, maybe because learners' proficiency in the L2 was not high enough for this to take place (e.g. de Groot et al., 2002 and, quite more recently, Krogh, 2022).

The CFE has also been found when words are not presented in isolation, but embedded in a sentence or paragraph. In this case, research has also been more extensive in the L2 than in the L1. However, evidence for the CFE in both languages has been found, for example, in the work of Van Assche and colleagues, who studied the CFE during sentence reading with eye-tracking in the L1 (Van Assche et al., 2009) and the L2 (Van Assche et al., 2011). Van Assche et al. (2009) asked Dutch-English bilinguals to read sentences in Dutch containing either a Dutch-English cognate or a non-cognate control word. Cognates were read significantly faster than non-cognates in terms of gaze durations and go-past times. Similarly, Van Assche et al. (2011) had the same participant profile read English sentences containing a Dutch-English cognate or a control word while their eye movements were monitored. Overall, cognates elicited shorter first-fixation durations, gaze durations and go-past times than controls[3]. Other studies supporting the CFE in a reading context are, in the L1, Cop et al. (2017) and in the L2, Balling (2013), Bultena et al. (2014), Duyck et al. (2007), Liebben and Titone (2009), Cop et al. (2017) and Van Assche et al. (2013, Experiment 2).

[3] First-fixation duration: the duration of the first (or only) fixation on a word when a sentence or text is first read. Gaze duration: the total amount of fixation time on a word until the eyes move to another word, either to the right or to the left of the target. Go-past time: the sum of all fixation time from the first fixation on a target word to —but not including— the first fixation on the word to its right.

Almost all studies exploring the CFE in word recognition and processing are conducted in the visual modality. Yet, there is some evidence that aurally presented cognates may also be recognized and processed faster and more accurately than non-cognates, both in the L1 and the L2. To name a couple of examples, Fricke (2022) conducted an online auditory English lexical decision experiment with English monolinguals, English-Spanish bilinguals and English natives having Spanish as a heritage language. Each participant performed the lexical decision task with English words (half cognate with Spanish, half non-cognate) and nonwords based on English words. Words were presented in the carrier sentence "Now I'll say *X*" and participants were instructed to decide whether the last word was a real word in English or not. Overall, accuracy was higher when recognizing cognates than non-cognates in all participant groups. Turning to the L2, Andras et al. (2022) had English monolinguals and Spanish-English bilinguals perform an auditory English word recognition experiment using eye-tracking and the visual world paradigm. Bilinguals were divided into two groups having either a low or a high proficiency in the L2. In each trial, participants saw a pair of pictures and subsequently listened to either a Spanish-English cognate differing in the degree of cross-linguistic phonological overlap (high or low) or a non-cognate. They had to select, as quickly as possible, the picture that matched the word heard. Reaction times and proportion of fixations on the target and the distractor pictures were calculated. Results showed a CFE in the low frequency group. On the one hand, these bilinguals were faster selecting the picture that matched cognate words than non-cognate words. On the other hand, their proportion of fixations on the target pictures was higher when these pictures matched cognates with low cross-linguistic phonological overlap than non-cognates, indicating that the first words were recognized faster than the second. Additional evidence for the CFE in spoken speech processing in the L1 and the L2 can be found in Blumenfeld and Marian (2007).

The influence of cognates on word processing within the Bilingual Interactive Activation plus (BIA+) model

As advanced, the most influential model of bilingual word processing capable of accounting for the CFE is the BIA+ model. The architecture of this model was detailed in Chapter 3 (Section 1.1) and will be briefly summarised here for convenience. The BIA+ model proposes that there is a single lexicon for the two languages of a bilingual and that lexical access is language non-selective. The processing system consists of orthographic, phonological and semantic word-form levels of representation linked to each other (for simplicity's sake, the authors do not include a syntactic level of representation, but this may be assumed, see Dijkstra & van Heuven, 2002). The model was originally designed to account for bilingual visual word recognition and comprehension, but Dijkstra and van Heuven argue that it could be adapted to account for auditory word recognition and processing. In short, the model proposes that when a word is visually presented, compatible L1 and L2 orthographic representations activate. Activation spreads from these orthographic representations to linked phonological and semantic representations. At the same time, activation spreads in a top-down manner from active semantic representations to linked phonological and orthographic representations. Active items at each level of representation compete with one another through mutual inhibition. Each orthographic and phonological word representation is linked to a language node responsible for inhibiting active representations in the non-target language to help the processing system select the appropriate word candidate. The most active item is the one selected. In what follows, I discuss the CFE within this model focusing on non-identical cognates, the type of cognates studied in this chapter (but see Peeters et al., 2013 for a discussion about the representation and processing of identical cognates within the BIA+ model).

Since cognates share a large number of orthographic features across languages, when a cognate is perceived, the orthographic representations of the cognate in the L1 and the L2 activate. By contrast, since non-cognates do not have a similar orthographic form across languages, when a non-cognate is perceived, the orthographic representation of the target word activates, but not that of its translation equivalent. In both cases, activation spreads in a bottom-up manner to associated phonological representations and to a semantic representation shared for translation equivalents. The activation of this semantic representation spreads back to linked phonological and orthographic representations. Crucially, while when a non-cognate is perceived the shared semantic representation receives activation from the orthographic and phonological representations of just the target word, when a cognate is perceived the shared semantic representation receives activation from the orthographic and phonological representation of the target word and its translation in the non-target language. Consequently, the semantic representation of cognates is more strongly activated than that of non-cognates. In addition, since the activation of the semantic representation feeds back to phonological and orthographic representations, the phonological and orthographic forms of cognates are also more strongly activated than those of non-cognates. The overall stronger activation of cognates compared to non-cognates facilitates their identification, for instance, in lexical decision and progressive demasking tasks and, more generally, causes them to be processed (e.g. read) faster than non-cognates. In the next section, I discuss how the explanation of the CFE within the BIA+ model could also be valid within the MOGUL framework. This was introduced in Chapter 1 and is further described below.

The influence of cognates on word processing within MOGUL

The MOGUL framework and the BIA+ model are similar in more than one aspect. MOGUL shares with the BIA+ the non-selective view of lexical access (Sharwood Smith & Truscott, 2014, pp. 249–250). It also shares with it the idea that there are orthographic, phonological, syntactic and semantic representations for each word, even if these representations are not part of an individual entity called *mental lexicon*, but each is found in a separate sublexicon in the visual, phonological, syntactic or conceptual modules, respectively. The set of representations constituting a word are coindexed and each representation is linked to the representation in adjacent modules by means of an interface (Sharwood Smith & Truscott, 2014, sec. 2.3.3). In MOGUL, each time that a representation activates, the items it is composed of also activate, together with other representations containing these items (Sharwood Smith & Truscott, 2014, sec. 3.3.5). Thus, when a word is visually and/or aurally presented, not only its orthographic and phonological representations activate (i.e. there is a rise in their current activation level), but representations of words sharing orthographic and/or phonological features with that word also activate. In line with the non-selective view of lexical access, this includes representations of words in the two languages of the bilingual. These representations then activate coindexed syntactic and semantic representations. Activation is bidirectional and, as such, it also spreads from semantic representations to coindexed syntactic, phonological and orthographic representations (Sharwood Smith & Truscott, 2014, sec. 3.4.4). Active representations compete to be selected by the processor in their module and the most active item is the one that wins the competition (Sharwood Smith & Truscott, 2014, sec. 3.4.1). Instead of language nodes, as in the BIA+ model, MOGUL proposes that acoustic and orthographic representations of words are connected to a conceptual representation of the language they belong to through an interface between the auditory/visual module and the conceptual module. These conceptual representations of languages, together with language-specific sounds and orthographic features, contribute to selectively activating the representations of the target language more strongly than those of the non-target language (Sharwood Smith & Truscott, 2014, sec. 6.4.3).

Importantly, it could be argued that, just as in the BIA+ model, MOGUL assumes that translation equivalents have a single semantic representation. This claim is based on how Sharwood Smith and Truscott propose that the meaning for a new L2 word is established (2014, sec. 7.6.1). The authors claim that, most often, the L2 word is presented together with its L1 translation or in a context where the L1 translation can be easily identified, either visually or aurally. In that case, the chain of orthographic and/or phonological, syntactic and conceptual representations of the L1 word would be active at the same time that the chain of representations of the L2 word is created. Since the L2 word would have never been encountered before, it would have to be linked to a conceptual representation. The most active one and, therefore, the one chosen, would be the conceptual representation of its L1 equivalent. Unless the context suggests that the meaning of the L2 item is significantly different from that of its L1 counterpart, the conceptual representation of the L1 word will remain that of the L2 word. This suggests that translation equivalents share a conceptual representation.

Taking all this into account, I propose that the MOGUL framework could support a representation and retrieval of cognates and non-cognates similar to the one in the BIA+ model and, hence, a similar explanation of the CFE. I argue that when a bilingual sees a cognate, the orthographic representation of that word activates and, due to its cross-linguistic similarity, the form of the cognate in the non-target language activates as well. Conversely, when a bilingual sees a non-cognate, its orthographic representation activates, but not that of its translation equivalent, for they share no orthographic features. Activation spreads from the orthographic representation(s) of the cognate or the non-cognate to coindexed phonological and syntactic representations in adjacent modules, as well as to the conceptual representation that translation equivalents share. As a result, the activation of the conceptual representation of the cognate will be stronger than that of the non-cognate. At the same time, the stronger activation of the cognate compared to the non-cognate's conceptual representation will spread to coindexed syntactic, phonological and orthographic representations. In sum, the chain of representations of the cognate will have a higher current activation level than that of the non-cognate. This will have an impact on the words' resting activation level. When processing finishes, representations having a high current activation level fall at a higher resting activation level than representations having a lower current activation level (Sharwood Smith & Truscott, 2014, sec. 3.3.5). Hence, since cognates have a higher current activation level than non-cognates, they will also have a higher resting activation level. A consequence of this is that the current activation level of cognates, defined as the sum of the resting activation level and any other activation received during the current processing, is higher than that of non-cognates not only due to cognates' cross-linguistic formal similarity, but also because their starting point of activation (their resting activation level) is higher.

The difference in current and resting activation level between cognates and non-cognates will have consequences for processing. First, each processor selects for processing the most active item in its module. Therefore, the items that have the highest current activation level are the ones selected (Sharwood Smith & Truscott, 2014, sec. 3.3.5). Since cognates have a higher current activation level than non-cognates, the former are more likely to be the most active words among competitors and, hence, to be selected for processing. Second, the higher the resting activation level of a representation is, the faster it becomes available for processing (Sharwood Smith & Truscott, 2014, sec. 3.3.5). Thus, the higher resting activation level of cognates compared to non-cognates would cause them to be available for selection by processors faster[4]. All this would explain why cognates are processed faster and more accurately than non-cognates.

1.2. The influence of cognates on syntax processing

As mentioned in Chapter 3, words are normally encountered in a sentence and, thus, must be recognized, processed and integrated into a syntactic structure that also needs to be processed. Chapter 4 studies whether and how cognates influence L2 syntax acquisition. To

[4] Note that the higher resting activation level of cognates compared to non-cognates must not be confused with the higher resting activation level of high frequency words compared to low frequency words proposed in Chapter 3, which was the result of the former being encountered in the input more often than the latter. Some authors (Lalor & Kirsner, 2001; Strijkers et al., 2010) have proposed that the CFE could be just a frequency effect. They argue that the fact that the representations of the cognate in the two languages of the bilingual activate each time that the L1 or the L2 word is encountered leads to an increase in the frequency of the cognate in the two languages. This reasoning is hard to reconcile with evidence that the processing of cognate words is affected by the corpus frequency that the cognate has in each of the bilingual's two languages, which suggests that the frequency of language-specific cognate representations is kept separate (e.g. Peeters et al., 2013).

this aim, it is important to comprehend first the influence that cognates might have on syntax processing. Lexical and syntactic processing have predominantly been explored individually and the interaction between the two has been largely overlooked. In spite of this, there is evidence that lexical and syntactic information interact throughout sentence processing and, crucially, that cognates may facilitate syntactic processing. The first account arguing in favour of the facilitative role of cognates was the Shared Syntax account (original account, Hartsuiker et al., 2004; account addressing the role of cognates, Hartsuiker & Pickering, 2008). This account proposes that similar structures are shared across languages and that processing such structures with a cognate might make them more available for subsequent production and comprehension. In the most recent version of the Shared Syntax account (Hartsuiker & Bernolet, 2017), syntactic representations become shared only as proficiency increases[5]. The claim that similar structures are shared across languages is based on evidence from bilingual syntactic priming, i.e. the fact that exposure to a structure in one of the languages of the bilingual favours the production of a sentence with that structure in the other language of the bilingual. This bilingual syntactic priming has been found from the L1 to the L2 (e.g. Cai et al., 2011; Hartsuiker et al., 2004; Kantola & van Gompel, 2011; Loebell & Bock, 2003; Schoonbaert et al., 2007) and from the L2 to the L1 (e.g. Cai et al., 2011; B. Chen et al., 2013; Loebell & Bock, 2003). Although the majority of studies have investigated priming in production, there is evidence that priming effects could generalise to comprehension. Specifically, previous research has found that exposure to a syntactic structure in the L1 favours the prediction and comprehension of a similar structure in the L2 (X. Chen et al., 2023; Hsieh, 2017).

The Shared Syntax account assumes that there is a shared lexicon for the two languages of the bilingual, in line with models of bilingual word processing such as the BIA+. It maintains that this lexicon consists of different levels of representation, namely the word-form level (including orthographic/phonological representations of words), the lemma level (including lemma nodes equivalent to a word's base form) and the conceptual level (including conceptual nodes representing the meaning of a word and the arguments it takes). The account assumes that each word form is linked to a lemma node and to a conceptual node that is shared for translation equivalents. Lemma nodes for each word are additionally linked to language nodes and to nodes specifying syntactic information such as word category or combinatorial properties, e.g. whether a verb is followed by a double object (DO) or by a prepositional object (PO). These nodes with combinatorial information (*combinatorial nodes*) are assumed to be linked to all relevant lemma nodes in the L1 and the L2. Priming is taken to be the consequence of shared combinatorial nodes retaining some activation after usage. For

[5] A debate has been held around what is required for two structures to be *similar* in the L1 and the L2. Some studies (e.g. Bernolet et al., 2007; Jacob et al., 2017; Kidd et al., 2015; Loebell & Bock, 2003) proposed, first, that different languages only shared syntactic representations for constructions that had the same word order. Nevertheless, further research showed that this was not true (Bernolet et al., 2009; B. Chen et al., 2013; Muylle et al., 2020; Shin & Christianson, 2009). Presence/absence of case marking was also found not to be a distinctive enough feature to consider two structures as different (e.g. Fleischer et al., 2012; Hartsuiker et al., 2016; Muylle et al., 2020). This led some authors to conclude that bilinguals share abstract syntactic representations.

instance, when an English native incrementally processes a sentence such as *The man offers the book to the woman*, the lemma node for the verb *offer* and the combinatorial node indicating that this verb can combine with a PO activate. This helps the speaker predict that the upcoming structure will include a PO. When sentence processing terminates, the PO node maintains some activation. If the speaker then needs to produce or process for comprehension an L2 sentence with a verb taking a DO or a PO, the lemma for that verb and the DO and PO nodes will activate. Since the PO node will retain some activation due to its prior use, the speaker will be most likely to produce a construction with a PO or predict that such a construction will be processed. This cross-linguistic priming effect in production and comprehension is obtained when the prime and the target sentences share no lexical items (e.g. Cai et al., 2011; Hartsuiker et al., 2004; Hsieh, 2017), but is boosted by the use of cognates in the prime and the target (Bernolet et al., 2012; Cai et al., 2011; X. Chen et al., 2023; J. Huang et al., 2019). Although I have focused on bilingual syntactic priming, it must be noted that priming also occurs within languages and that it is boosted by the use of the same word in the prime and the target sentences (see, for instance, Arai et al., 2007; Pickering & Branigan, 1998). In what follows, I review a couple of studies evidencing the lexical boost and, most importantly, the cognate boost to priming in production and comprehension.

In the production domain, Cai et al. (2011) had Cantonese natives with a very high proficiency in L2 Mandarin perform two experiments evaluating between-language and within-language priming. Experiment 1 studied priming from Mandarin to Mandarin (L2-L2) and from Cantonese to Mandarin (L1-L2). Experiment 2 studied priming from Cantonese to Cantonese (L1-L1) and from Mandarin to Cantonese (L2-L1). The procedure was the same in the two experiments. Participants listened to a prime sentence, which could be a construction in Mandarin or in Cantonese including either a DO or a PO, formed similarly in the two languages (i.e. in the PO construction the theme is followed by the recipient, which is preceded by the preposition *gei* (Mandarin)/*bei* (Cantonese), equivalent to the English “to”. The order of arguments is the inverse in the DO construction). Next, participants saw a prime picture, which matched the prime sentence on half of the occasions, and had to decide whether the picture was an appropriate representation of the sentence or not. Then, the prime picture was replaced by the target picture, together with a sentence beginning in Mandarin (Experiment 1) or Cantonese (Experiment 2) which promoted that participants produced constructions with a DO or a PO (instead of other constructions). These sentences had to be produced with a verb having the same meaning as the one in the prime sentence (the exact same verb in within-language trials, a cognate in between-language trials) or having a different meaning. Experiments 1 and 2 provided evidence of structural priming (there was a general tendency to use the structure in the prime sentence to describe the target picture). Priming was greater when the verb in the prime and the target sentence had the same as opposed to a different meaning, both within languages (lexical boost to priming) and between languages (cognate boost to priming).

A recent study by X. Chen et al. (2023) provided evidence of a cognate boost to priming in comprehension. The authors conducted two experiments with Cantonese-Mandarin-English trilinguals with a high proficiency in the L2 and an intermediate proficiency in the L3. The two experiments investigated within-language priming (from Mandarin to Mandarin, L2-L2) and between-language priming (from Cantonese to Mandarin, L1-L2, and from English to Mandarin, L3-L2) using PO and DO constructions, formed similarly in the three languages. In Experiment 1, the verbs in the prime and the target sentences had different meanings. In Experiment 2, the verbs in the prime and the target sentences could have either different or similar meanings. That is, they were either non-translation equivalents or (i) identical verbs (when the prime and the target sentences were in Mandarin), (ii) cognates (when the prime sentence was in Cantonese and the target in Mandarin) or (iii) non-cognates (when the prime sentence was in English and the target in Mandarin). The two experiments used a visual world eye-tracking paradigm. In each trial, participants read aloud either a DO or a PO sentence in Cantonese, Mandarin or English (Cantonese and Mandarin have the same orthographic system, different from English). Next, a picture of an action involving an agent, a theme and a recipient appeared. After 1500ms, a Mandarin sentence played (either a PO or a DO construction) and eye-movements to each of the three entities in the picture were recorded. Crucially, the PO and DO structures were temporarily ambiguous from the onset of the sentence until the second syllable of the first NP (e.g., ***Yeye huan Qiu****yuan yifu Qiupai*, lit. “Grandpa returns football player a racket” (DO) or ***Yeye huan Qiu****pai gei Qiuyuan*, lit. “Grandpa returns racket to football player” (PO)). X. Chen et al. predicted that after listening to a prime DO sentence the DO combinatorial node would retain some activation, which would facilitate its reactivation during target sentence processing (in the same language and across languages). Consequently, participants would predict the temporarily ambiguous sentence to be a DO, as indexed by a significantly larger proportion of looks to the recipient than to the theme during the temporary ambiguity. The opposite should hold after listening to a prime PO sentence. This prediction was met when the prime and the target sentences were in Mandarin (within-language priming, Experiments 1 and 2) and when the prime sentences were in Cantonese and the target sentences were in Mandarin, buy only when they contained cognate verbs (between-language priming, Experiment 2).

Several explanations for the cognate boost to priming have been proposed. On the one hand, it has been argued that when a structure with a cognate is processed, the orthographic/phonological form of the cognate in the target and the non-target language activates. The two forms spread their activation to the lemma level of representation, activating the lemmas of the cognate in the two languages. Because these lemmas are simultaneously activated, a link develops between them. This link does not develop for non-cognate lemmas in the L1 and the L2, since each time that a structure with a non-cognate is processed, only the non-cognate word form and its lemma in the target language activate. The lemma link causes that a stronger activation spreads from the lemma level of representation to the shared combinatorial node each time that a cognate is processed and, thus, that the residual activation of this combinatorial node is stronger after processing a sentence with a cognate than a non-cognate. This facilitates subsequent processing of a similar structure with a cognate in the bilingual's other language, yielding the cognate boost to priming observed (J. Huang et al., 2019). On the other hand, an alternative explanation might do without the lemma link. It might be that the activation of the orthographic/phonological form of the cognate in the two languages of the bilingual causes that the two cognate lemmas activate and that these feed stronger activation to the shared combinatorial node. Activation would also spread from the two cognate lemmas to the shared conceptual representation, causing its activation to be stronger and, thus, to feed stronger activation back to the lemmas and the shared combinatorial node. All this would make the combinatorial node more available for subsequent processing when a structure is processed with a cognate than with a non-cognate (X. Chen et al., 2023). This explanation would also be valid in a model of bilingual sentence production or comprehension that does not include a lemma level of representation, but which believes that there is a direct link between word forms and combinatorial nodes (Bernolet et al., 2012).

A second approach to the facilitative role of cognates in syntax processing is the Lexical Bottleneck Hypothesis (Hopp, 2018). This hypothesis has mostly accounted for cognate effects when processing cross-linguistically dissimilar structures and, more precisely, when processing sentences in which an L1 structure is in competition with a different L2 structure. The Lexical Bottleneck Hypothesis assumes that lexical processing occurs before syntactic processing and directly influences it. Being a hypothesis, not a model or a theory, it does not specify exactly how the mapping from lexical to syntactic processing occurs, but it assumes that some degree of seriality is necessary. Regarding the influence of lexical processing on syntactic processing, the hypothesis proposes that a costly (e.g. slower) lexical processing may exhaust all the resources necessary to perform a native-like syntactic computation. Accordingly, it makes two predictions. First, that "delays or differences in earlier stages of processing that subserve syntactic processing may lead to non-target syntactic processing" (2018, p. 17). Second, that "if difficulties, delays or cross-linguistic influence in lower-level processing are removed or taken into account, adult L2 learners can come to demonstrate target-like syntactic processing in the L2" (2018, pp. 17–18).

Unlike the Shared Syntax account, which addresses whether processing structures with a cognate makes them more available for subsequent processing, the Lexical Bottleneck Hypothesis is concerned with how cognates affect the current processing event. Hopp postulated the hypothesis based on the findings of several L2 syntactic processing studies manipulating the lexical processing cost of some of the words embedded in the target structures, using, amongst others, cognate words. To name an example, Hopp (2017) conducted two eye-tracking experiments investigating how cognates and non-cognates modulated syntactic co-activation during L2 sentence comprehension. German natives with an intermediate or advanced proficiency in L2 English read sentences containing either reduced or full relative clauses in initial position. The two types of relative clauses appeared in embedded clauses and in main clauses. Thus, there were four sentence types (1a. Reduced relative clause in embedded clause, 1b. Reduced relative clause in main clause, 1c. Full relative clause in embedded clause and 1d. Full relative clause in main clause. The embedded or main clauses are in brackets, the (reduced) relative clauses are in italics):

(1) a. [When the doctor *Sarah ignored* tried to leave the room] the nurse came in all of a sudden.
b. [The doctor *Sarah ignored* tried to leave the room] when the nurse came in all of a sudden.
c. [When the doctor *who Sarah ignored* tried to leave the room] the nurse came in all of a sudden.
d. [The doctor *who Sarah ignored* tried to leave the room] when the nurse came in all of a sudden.

(Hopp, 2017, p. 105)

Participants read the four sentence types for comprehension, as well as fillers exclusively in English (Experiment 1) or alternating between English and German (Experiment 2). Reduced relative clauses differed in whether they could be temporarily parsed according to German syntax or not. In German, the canonical word order in embedded clauses is SOV but, in main clauses, the verb always occupies the second position. Consequently, German natives could temporarily activate the L1 SOV word order when reading embedded reduced relative clauses such as (1a), but not when reading main reduced relative clauses such as (1b). Hopp predicted that if learners activated L1 syntax during L2 reading, reading times for embedded reduced relative clauses would be longer than for main reduced relative clauses. That is, in the first case, the interference from the L1 word order would cause learners to misanalyse the embedded clauses, which would then need to be reanalysed according to the L2 word order. Full relative clauses (1c and 1d) acted as control sentences accounting for any differences in the processing of main and embedded clauses that were not motivated by syntactic co-activation. Importantly, half of the verbs in the relative clauses were cognates in German and English and the other half were non-cognates. Hopp hypothesized that retrieving cognates from the mental lexicon would be easier than retrieving non-cognates and, as a result, that in the first case learners would dispose of more resources to inhibit the L1 syntax and to compute an L2 target-like parse. If this were the case, embedded reduced relative clauses with non-cognates would be read slower than the other types of relative clauses. In Experiment 1, intermediate-proficiency learners showed slower first-pass reading times for verbs in embedded reduced relative clauses than for verbs in the other relative clauses, indicative of the fact that, in the former, learners computed an L1-based parse prior to reanalysing the structure in terms of L2 grammar. Crucially, the reading slowdown was limited to non-cognate verbs. When the verb in the embedded reduced relative clause was cognate in the L1 and the L2, reading times showed no sign of interference form an L1-based parse. The results of intermediate-proficient learners extended to the whole group of learners when German fillers boosted co-activation of the L1 in Experiment 2.

After the formulation of the Lexical Bottleneck Hypothesis, a couple of studies by Soares and colleagues tested its prediction of a facilitative role of cognates in syntax processing. In the first one, Soares et al. (2018) explored how including cognates vs. non-cognates in English ambiguous relative clauses affected the attachment strategies adopted to resolve the ambiguity (high attachment (HA) or low attachment (LA)). Participants were L1 European Portuguese speakers with an intermediate or an advanced proficiency in L2 English. European Portuguese natives have a preference for HA, whereas English natives prefer LA (Cuetos & Mitchell, 1988; Soares et al., 2010). Participants performed a relative clause completion task in English. They were asked to read the beginning of relative clauses and complete them as quickly as possible with the first continuation they could think of. A complex NP containing two nouns, either cognate or non-cognate in Portuguese and English, always preceded the relative clause. Accordingly, there were four experimental conditions (2a. Cognate noun – Cognate noun (C-C), 2b. Cognate noun – Non-cognate noun (C-NC), 2c. Non-cognate noun – Cognate noun (NC-C) and 2d. Non-cognate noun – Non-cognate noun (NC-NC). The Portuguese translation of each sentence is between brackets and cognates and non-cognates are underlined). All relative clauses could be completed according to a HA or LA strategy, depending on whether the clause was associated to the first or the second noun of the complex NP.

(2) a. Britney recognized the guard of the prisoner who…
[A Beatriz reconheceu o guarda do prisioneiro que…].
b. Bessie had tea with the fan of the singer who…
[A Bruna foi tomar chá com a fã da cantora que…].
c. The shopkeepers saw the thieves of the tourists who…
[Os comerciantes viram os ladrões dos turistas que…].
d. Molly loved the box of the cake that was…
[A Maria adorou a caixa do bolo que…].
(Soares et al., 2018, pp. 172–173)

Overall, learners showed a target-like LA preference to disambiguate the relative clauses. Specifically, both intermediate and advanced L2 learners completed a below-chance number of sentences using a HA (L1) strategy in all conditions but the NC-C, in which HA completions were at chance. In addition, cognates seemed to facilitate a target-like ambiguity resolution, for learners produced more LA completions in the C-C condition than in the NC-NC condition. The Lexical Bottleneck Hypothesis would explain this by claiming that the less demanding lexical processing of two cognates compared to two non-cognates freed the resources necessary to inhibit the L1 attachment preference. While this study provided additional evidence in favour of the facilitative role of cognates in L2 syntax processing, it used an offline task and, as such, it measured the result of syntactic processing with cognates vs. non-cognates, not how real-time syntactic processing was affected by cognates.

To investigate this, Soares et al. (2019) conducted a follow-up experiment. They studied how European Portuguese natives with intermediate or advanced proficiency in L2 English resolved the HA-LA ambiguity with the same type of relative clauses and lexical manipulation. Yet, this time they monitored participants' eye movements during sentence reading. As in the 2018 study, a complex noun phrase with two nouns preceded the relative clause. These nouns could be both cognate (C-C condition), both non-cognate (NC-NC condition), the first cognate and the second non-cognate (C-NC condition) or the first non-cognate and the second cognate (NC-C condition). Relative clauses were temporarily ambiguous between HA and LA until a critical word, which bore a semantic relation with one of the two nouns in the complex NP and forced the sentence to be interpreted according to a HA or LA strategy (e.g. *Britney recognized the guard of the prisoner who had been honoured for his braveness*, HA strategy). Participants read sentences for comprehension. Half of the sentences in each condition were disambiguated using HA and, the other half, using LA.

Soares and colleagues argued that, if as predicted by the Lexical Bottleneck Hypothesis the less costly processing of cognates compared to non-cognates freed more resources to inhibit L1 syntax, overall reading times for the disambiguating word would be longer in the NC-NC than in the C-C condition. In addition, because L1 syntax would be easy to inhibit when the relative clause was preceded by cognates, in the C-C condition reading times for the disambiguating word in LA sentences would be shorter than in HA sentences. By contrast, because L1 syntax would be harder to inhibit when the relative clause was preceded by non-cognates, in the NC-NC condition the HA and LA strategies would be comparably available for processing, leading to comparable reading times for the disambiguating word in HA and LA sentences. Finally, if there was any L1 syntax interference in sentences disambiguated with a LA strategy, this would be the least noticeable in the C-C condition, i.e. reading times in this condition would be shorter than in the rest. The results matched these predictions of the Lexical Bottleneck Hypothesis.

In sum, in this section I reviewed studies framed within the Shared Syntax account that showed that processing a sentence with a cognate in one of the languages of the bilingual facilitates production and comprehension of a sentence with a similar structure in the other language of the bilingual. Additionally, I reviewed studies supporting the Lexical Bottleneck Hypothesis that revealed that when a sentence is temporarily ambiguous between an L1 structure and a different L2 structure, embedding cognates in that sentence might facilitate a target-like L2 parse. Both the Shared Syntax account and the Lexical Bottleneck Hypothesis aim to explain a reduced set of observed phenomena; they do not address how sentences are processed in real time nor how exactly words are integrated into the structure being processed. If cognates affect L2 syntax processing, this would arguably be better accounted for by a language-processing model or framework that explains the cognate facilitation effect, describes real-time sentence processing and the interplay between the two. I believe that an appropriate framework could be the MOGUL. In what follows, I describe how I propose that differences in cognateness could influence processing of an unambiguous L2 syntactic structure within this framework, for these are the type of structures studied in this chapter. Putting cognateness aside, the description of how sentence processing takes place will be based on the one for *kick the ball* in Chapter 3 (Section 1.2), which was directly taken from MOGUL (Sharwood Smith & Truscott, 2014, sec. 4.5.3).

The influence of cognates on syntax processing within MOGUL

I will consider how a Spanish-English bilingual would process the English sentences in (3), with an underlying syntactic structure such as $[_{CP}$ NP $[_{VP}$ V NP]]. The verb in (3a), *buy,* is non-cognate with its Spanish translation, *comprar.* The verb in (3b), *admire,* is cognate with its Spanish translation, *admirar.* For simplicity's sake, I will imagine that *buy* and *admire* have a similar frequency of use (even if this is not the case in English) so that frequency is not a variable influencing processing. Additionally, I will focus on spoken speech processing, leaving visual processing aside.

(3) a. Mary buys a painting.
b. Mary admires a painting.

When processing (3a), the bilingual would first listen to the word *Mary* and the phonological representation of the word would activate in the phonological store. Activation would spread from this representation to the coindexed syntactic representation in the syntactic store, e.g. $[N_j]$ and to other representations containing it, possibly including $[_{CP}$ $[_{NP}$ $N_j]$ $[_{VP}$ V NP]]. Next, the bilingual would hear the verb *buys.* A phonological representation would activate, spreading its activation to the coindexed syntactic representation, e.g. $[V_i]$. The activation of this representation would spread to the verb's subcategorization frame, $[V_i$ $NP_{p,\, q,\, r\ldots}]$, and to larger syntactic representations including $[V_i]$, amongst which the previously activated $[_{CP}$ NP $[_{VP}$ V_i NP]]. At the same time, activation would weakly spread from $[V_i$ $NP_{p,\, q,\, r\ldots}]$ to NPs used with the verb *buy*, maybe including *a painting*. Meanwhile, *a painting* would be processed. Its phonological representation would activate, as well as its coindexed syntactic representation, e.g. $[_{NP}$ D_x $N_b]$. Activation would spread from this NP to other structures containing it, increasing the activation of $[_{CP}$ NP $[_{VP}$ V_i NP]] and causing the syntactic processor to select it amongst any alternatives. The NP (*a painting*) would be combined with $[V_i]$ (*buys*) to form a VP (*buys a painting*), which would then be merged with the preceding NP (*Mary*) to form a CP (*Mary buys a painting*).

Processing of (3b) would be essentially as described for (3a), with the difference that in this case the verb is cognate between English and Spanish. To comprehend the influence that differences in cognate status would have on syntactic processing, it is necessary to remember how cognate and non-cognate words would be processed within MOGUL (cf. Section 1.1). I proposed that when a bilingual encounters a cognate, its phonological form in the bilingual's two languages activates due to the formal similarity between the two. These phonological representations spread their activation to coindexed syntactic representations and to a conceptual representation shared between the L1 and the L2 translation equivalents. By contrast, when a bilingual encounters a non-cognate, its phonological form in the target language activates, but not that in the non-target language. This phonological representation spreads its activation to coindexed syntactic and conceptual representations. Crucially, the conceptual representation of the cognate will be more strongly activated than that of the non-cognate, since it will receive activation from two word forms, and it will feed back stronger activation to coindexed syntactic and phonological representations. Hence, I assume that the syntactic representation of the cognate verb in (3b) will be more activated than that of the non-cognate verb in (3a).

As described, MOGUL assumes that the syntactic representations of lexical items spread their activation to larger syntactic representations containing them, increasing their current activation level (in 3a and 3b, the syntactic representation of the verb would spread its activation to [$_{CP}$ NP [$_{VP}$ V NP]]). The stronger activation spreading from the syntactic representation of the cognate verb compared to the non-cognate verb would cause that the current activation level of the structure containing the cognate is higher than that of the structure containing the non-cognate. As mentioned, representations with a higher current activation level fall at a higher resting activation level when processing terminates and, the higher the resting activation level of a representation is, the faster it becomes available for processing (Sharwood Smith & Truscott, 2014, sec. 3.3.5). Thus, I propose that the structure processed with a cognate (3b) will have a higher resting activation level and will be more readily available for subsequent processing than the structure processed with a non-cognate (3a). To some extent, this explanation resonates with the one proposed within the Shared Syntax account, in the sense that the stronger activation of cognates compared to non-cognates spreads to syntactic representations linked to these words, making them more available for processing[6]. Overall, the reasoning in this section, based on activation, is very similar to the one presented in Chapter 3 for the effect of lexical frequency on syntax processing.

2. Cognates and the acquisition of cross-linguistically dissimilar L2 structures

This chapter investigates whether and how cognates facilitate L2 syntax acquisition by complete beginner adult learners. As advanced, to the best of my knowledge no study has assessed whether processing syntactic structures with cognate vs. non-cognate words influences learning of these structures. In the MOGUL framework, acquisition is the result of processing and it involves no learning-specific mechanisms (Sharwood Smith & Truscott, 2014, sec. 4.2). In this light, I propose that if differences in cognateness influence L2 syntax processing, as suggested in the previous section, these differences could also influence L2 syntax acquisition by processing. In this section, I will detail how exactly I propose this

[6] Note, however, the differences between the two: MOGUL does not assume a lemma level of representation; it argues that word forms are directly linked to syntactic representations via interfaces. Likewise, it does not assume that combinatorial nodes indicate the type of structure under construction, but it assumes that structures are stored as representations in the syntax module (Sharwood Smith & Truscott, 2014, sec. 2.3).

influence would play out, focusing on how cognates would affect the initial acquisition of the cross-linguistically dissimilar L2 structures studied in Experiment 4.

2.1. Cross-linguistically dissimilar structures studied in Experiment 4

The object of study of Experiment 4 are two Basque-based structures differing in word order and agent-patient marking with Spanish. While the agent-patient marking is the same in the two structures, the word order varies between SOV (4a) and OSV (4b). In these structures, the subject is always an agent (A) and the object is always a patient (P).

(4) a. Antzezle-ak gidari-a pintatu.
actor-A pilot-P paint
"The actor is painting the pilot."

b. Gidari-a antzezle-ak pintatu.
pilot-P actor-A paint
"The actor is painting the pilot."

(Example sentences used in Experiment 4)

The sentences in (4) consist of two animate nouns followed by a verb and express the same meaning. In (4a), the first noun is the agent and the second noun is the patient of the action denoted by the verb, as marked by the suffixes or postpositions *–ak* and *–a,* respectively. In (4b), the agent and patient marks are the same, but the order of the nouns is reversed. Thus, the two animate nouns *antzezle* ("actor") and *gidari* ("pilot") could potentially be the agent or the patient of the verb *pintatu* ("paint"). The only way to differentiate between the agent and the patient in the reversible sentences in (4) is by means of the agent and patient marks. This word order and agent-patient marking resemble the ones in Basque, although some differences must be noted. Consider, for instance, the Basque SOV and OSV sentences in (5a) and (5b). As shown, SOV and OSV sentences may consist of two nouns followed by a verb. The verb *ikusi* ("see") is in the perfect participle (PTCP). This is followed by an auxiliary verb (*du,* AUX) agreeing with the subject and the object in person and number (in both cases, third person singular, 3SG) to construct a sentence in the present perfect tense. The verb in the structures studied in Experiment 4 is also a perfect participle in Basque but, as shown in (4), the structures did not include the auxiliary. This is because participants (Spanish natives without knowledge of Basque) learnt the verbs in their citation form, which is the participle form (cf. Procedure, Section 3.4). Including the auxiliary in the structures would have made them unnecessarily more complex. In the sentences in Experiment 4, this participle verb is taken to express the meaning of a conjugated verb.

(5) a. Emakume-a-k gizon-a ikusi du gaur.
woman-D-ERG.A man-D.[ABS.P] see $_{PTCP}$ 3SG.AUX.3SG today
"The woman has seen the man today."

b.	Gizon-a	emakume-a-k	ikusi	du	gaur.
	man-D.[ABS.P]	woman-D-ERG.A	see $_{\text{PTCP}}$	3SG.AUX.3SG	today

“The woman has seen the man today.”

(Erdocia et al., 2009, p. 3)

The meaning of the suffixes –*ak* and –*a* in (4) is also different from the meaning these endings have in Basque. In Basque, when the head of a noun phrase is a common noun, it requires a determiner, which follows the noun (Laka, 1996). Going back to (5), in these sentences there are two noun phrases (*emakumeak,* “the woman” and *gizona,* “the man”) headed by two common nouns (*emakume,* “woman” and *gizon,* “man”). The determiner (D) that follows these nouns is the definite singular article, which corresponds to the suffix –*a*. In addition, Basque is a case-marking, ergative-absolutive language (Hualde & Ortiz de Urbina, 2003; De Rijk, 2007). Ergative-absolutive languages treat the argument of intransitive verbs as the object of transitive verbs, but have a special form for transitive subjects. Specifically, the subject of intransitive verbs and the object of transitive verbs are morphologically unmarked for absolutive case (i.e. they bear a covert case mark). By contrast, the subject of transitive verbs bears the ergative case mark –*k*. In (5), the noun phrase *gizona* (“the man”), which is the direct object and the patient of the transitive verb *ikusi du* (“has seen”), is in the absolutive case (ABS) and, thus, ends with the definite article –*a*. In addition, *emakumeak* (“the woman”) is the subject and agent of the transitive verb and bears the ergative (ERG) mark –*k*, which blends with the definite article resulting in an NP ending in –*ak*. In sum, while in Basque the agent and the patient of transitive verbs bear a –*k* and a zero morpheme, respectively, in Experiment 4 they ended in –*ak* and –*a*. This is because, as advanced, participants learnt nouns in their citation form (i.e. without the article or case mark) but, mirroring Basque grammar, the article and/or case mark were affixed to nouns in SOV and OSV sentences. Since participants had no previous knowledge of Basque, to them the endings –*ak* and –*a* marked the agent (grammatical subject) and the patient (grammatical object) of the sentence, respectively. In this light, I will henceforth refer to these suffixes as *agent-patient marks*.

Importantly, the word order and agent-patient marks studied differ from the word order and agent-patient marks in Spanish, participants’ native language. On the one hand, Spanish is head-initial and its canonical word order is SVO (López, 1997). Although in Spanish nearly all constituent combinations are possible, including SOV and OSV, these are extremely infrequent and are perceived by native speakers as marked (Carreiras et al., 1995). For instance, of all transitive sentences in the ADESSE corpus (García-Miguel et al., 2010), just 0.6% are SOV or OSV[7]. On the other hand, Spanish is a nominative-accusative language and treats the argument of intransitive verbs and the subject of transitive verbs equally, without overtly marking them for nominative case (NOM), but sometimes overtly marks the object of transitive verbs for accusative case (ACC). Specifically, when direct objects are animate,

[7] This corpus of Spanish contains 1.5 million words, 159,000 clauses and 3,450 verb lemmas, all manually annotated with syntactic and semantic information. These items are taken from different types of documents, including spoken texts, written press, narrative, essay and theatre.

specific (i.e. they point at an entity which the speaker has some knowledge of) and refer to particular individuals within a certain class, they are preceded by the preposition (PREP) *a* ("to" in English), a phenomenon known as *Differential Object Marking.* This preposition indicates that the noun that follows it is the direct object and the patient of the action denoted by the verb (Fábregas, 2013). An example of this can be found in (6), where *paciente* ("patient"), the direct object and patient of the transitive verb *salvó* ("saved", past third person singular form, PST.3SG), is preceded by *al*, the contraction of the preposition *a* and the definite masculine determiner *el* ("the").

(6) La enfermera salvó al paciente.
The nurse.NOM.A save $_{PST.3SG}$ PREP.D.P.ACC patient
"The nurse saved the patient."

(Gutiérrez, 2008, p. 370)

2.2. The influence of cognates on the acquisition of cross-linguistically dissimilar L2 structures

In this section, I hypothesize how embedding Spanish-Basque cognate verbs vs. non-cognate verbs in the cross-linguistically dissimilar structures presented in the previous section could affect their acquisition by Spanish natives without knowledge of Basque. I will mainly focus on acquisition by processing occurring in MOGUL's syntax module, mentioning the interaction with the acoustic, visual, phonological and conceptual modules when necessary. However, it should be remembered that in MOGUL processing occurs in all modules simultaneously as input is incrementally perceived and that, for each sentence, a chain of acoustic and/or visual, phonological, syntactic and conceptual representations is created (Sharwood Smith & Truscott, 2014, sec. 3.4.1). First, I will focus on how Spanish natives would process the SOV sentences in (7) from the moment they encounter them for the first time, assuming that they have already learnt the words in these sentences, as in Experiment 4 (see Section 3.4 for the details). In Basque, some concepts can be expressed using two words, one cognate with Spanish, as the verb *pintatu* in (7a) (in Spanish *pintar,* "paint") and one non-cognate with Spanish, as the verb *margotu* in (7b) (also translating into Spanish *pintar,* "paint")[8].

(7) a. Antzezle-ak gidari-a pintatu.
actor-A pilot-P paint
"The actor is painting the pilot."

b. Antzezle-ak gidari-a margotu.
actor-A pilot-P paint
"The actor is painting the pilot."

[8] These verbs are in the perfective participle form but, following Hualde and Ortiz de Urbina (2003), I take them to be equivalent to the infinitive form in Spanish because the participle is the unmarked or citation form of verbs in Basque.

To begin with, Spanish natives will encounter *antzezleak.* This is a compositional form, consisting of a noun (*antzezle*) and a suffix (*–ak*). Sharwood Smith and Truscott claim that when a compositional word is encountered, a whole form and a compositional form activate and compete for selection by the linguistic processors. The form that is more activated is the one selected for processing. The strength with which a compositional form is activated depends on its transparency, i.e. the facility with which the orthographic, phonological, syntactic and conceptual representations of the stem and affixes can be identified, and its frequency of occurrence, i.e. the more times a compositional form is encountered, the more opportunities to process it compositionally (Sharwood Smith & Truscott, 2014, sec. 4.5.1). Considering this, it could be argued that when Spanish natives read or listen to *antzezleak* for the first time, a chain of orthographic, phonological, syntactic and conceptual whole-form and compositional-form representations will activate and compete for selection by the processors in each module. On the one hand, learners will have learnt the noun in its citation form (*antzezle*), which may increase the transparency of the compositional form at the orthographic and phonological level, i.e. it may make it more salient. Nevertheless, while the orthographic and phonological representations of *antzezle* will activate the syntactic representation of a noun and the conceptual representation of ACTOR, there will be no syntactic and conceptual representations associated with *–ak* yet, and what these representations are will not be obvious. The non-transparency of the compositional form at the syntactic and conceptual level will probably cause that the whole-form representation of *antzezleak* is the one selected for processing.

The syntactic representation of the whole form (e.g. [N_y]) will extend its activation to syntactic representations including it, which will try to impose themselves on subsequent input as it is incrementally received. The first time that the sentence is processed, these representations will be those of the L1 (Spanish), possibly including [$_{CP}$ [$_{NP}$ N_y] [$_{VP}$ V NP]]. Additionally, in Spanish singular countable nouns are preceded by a determiner so, upon processing the noun, the representation [$_{NP}$ D N_y] will probably activate as well. I propose that a null orthographic and/or phonological representation may be coindexed with the syntactic representation of the determiner in the syntax module, i.e. learners might interpret that Basque allows determiners not to be overtly realized. Similarly, Sharwood Smith and Truscott propose that the syntactic representation of a determiner might be coindexed with a null phonological representation in English learners' linguistic system, since English requires an article before nouns, but many English NPs are not preceded by an overt determiner (Sharwood Smith & Truscott, 2014, sec. 10.4.2). This representation will increase in resting activation level each time that an NP without a determiner is processed. Meanwhile, learners will read or listen to *gidaria.* Processing of this word should be as described for *antzezleak,* with a whole-form chain of representations winning competition against a compositional-form chain of representations[9]. When the syntactic representation of this noun is activated, a mismatch will

[9] In Experiment 4, input was both visually and aurally presented, ruling out the possibility that learners parsed the agent mark on the first noun as a determiner preceding the second noun ([$_{CP}$ antzezle [$_{VP}$ [$_{NP}$ ak gidaria] pintatu]]).

be detected between the input and active syntactic representations in the syntax module (e.g. [$_{CP}$ NP [$_{VP}$ V NP]] from L1 Spanish).

Next, the verb will be processed. For the moment, I intentionally do not distinguish between cognates and non-cognates. The syntactic representation of the verb (e.g. [V_z]) will activate and the syntactic processor will merge it with the preceding NPs to create a syntactic representation of the sentence (e.g. [$_{CP}$ NP [$_{VP}$ NP V_z]]), which will initially receive a low resting activation level. By doing this, the subcategorization frame for this verb will be created (Sharwood Smith & Truscott, 2014, sec. 4.3.1). To interpret the sentence, the verb's arguments will be assigned a conceptual role. In MOGUL, these are conceptual representations in the conceptual module (Sharwood Smith & Truscott, 2014, sec. 2.5.1). Natural languages have a strong tendency to put agents before other arguments (the so-called *agent-first preference*), as shown by the predominance of word orders in which the subject of the transitive verb (semantically, the agent) is placed before its object (semantically, the patient). Specifically, 80% of the canonical word orders in the world's languages are SVO, SOV or VSO (Dryer, 2013). This includes Spanish, participants' L1 (SVO, López, 1997). As a result, Spanish natives prefer to interpret the first animate NP in a sentence as the agent rather than the patient, as shown by behavioural, electrophysiological and eye-tracking data (Gómez-Vidal et al., 2022; Zawiszewski et al., 2022). Considering this, I assume that Spanish natives will interpret the first NP (*antzezleak*) as the agent and the second NP (*gidaria*) as the patient. In other words, the syntactic representations of the first and the second NPs will be coindexed with the conceptual representations of AGENT and PATIENT, respectively. This will lead to a correct interpretation of the sentence.

When Spanish natives encounter an SOV sentence again, the newly created, verb-final syntactic representation will activate and will be used to process the input. If only SOV sentences were processed, learning of agent-patient marking would not be necessary, since the aforementioned syntactic structure, with a whole-form representation of the first NP coindexed with AGENT and a whole-form representation of the second NP coindexed with PATIENT, would be appropriate for processing. Nevertheless, in Experiment 4 learners processed not only SOV sentences, but also OSV sentences. I argue that this would trigger learning of agent-patient marking and, consequently, of the two target syntactic structures. In what follows, I describe how Spanish natives with no knowledge of Basque would process the equivalent OSV sentences in (8) from first exposure. Like in (7), in (8a) the verb is cognate between Spanish and Basque and in (8b) it is non-cognate.

(8) a. Gidari-a antzezle-ak pintatu.
pilot-P actor-A paint
"The actor is painting the pilot."

b. Gidari-a antzezle-ak margotu.
pilot-P actor-A paint
"The actor is painting the pilot."

I argue that the first time that OSV sentences are encountered, processing will be the same as for their SOV counterparts. Learners will construct a syntactic representation along the lines of [$_{CP}$ NP [$_{VP}$ NP V]] and, due to the agent-first preference, they will coindex the first whole-form NP (*gidaria*) with the conceptual role of AGENT and the second whole-form NP (*antzezleak*) with the conceptual role of PATIENT. The only difference between SOV and OSV sentences is that, in the second case, the conceptual representation of the sentence will not match the non-linguistic contextual information accompanying it (in Experiment 4, this information was provided by a picture representing the meaning of the sentence, see Section 3.4). That is, the interpretation of the sentence will be "the pilot is painting the actor" but the contextual information will indicate that it should be "the actor is painting the pilot". After misprocessing one or more OSV sentences, learners will realize that this language has flexible word order. For the sentences in (8), the non-linguistic context will lead to the activation of a conceptual representation of the sentence that coindexes ANTZEZLE and AGENT, and GIDARI and PATIENT. This conceptual representation will need to be coindexed, in turn, with an appropriate representation in the visual/phonological and syntactic modules. Since word order will have proved not to be a reliable cue to agentivity, the linguistic system will need to look for a different cue. I argue that this will lead to an increase in the current activation level of the compositional representations of *antzezleak* and *gidaria*, and that the conceptual roles of AGENT and PATIENT will be coindexed with the chain of representations of *–ak* and *–a.* Since the compositional representation of these NPs will prove useful for processing, when processing terminates the resting activation level of these representations will increase, and so will that of the structure containing them. Each time that an OSV sentence is subsequently processed, the resting activation level of the compositional form of the NPs will rise and they will eventually dominate over competing whole-form representations (Sharwood Smith & Truscott, 2014, sec. 4.5.1). This will also cause that, when processing an SOV sentence, the compositional representations are accessed and used. In sum, the resting activation level of the SOV and OSV structures with these compositional NP representations will increase after each processing and they will gradually become more robustly established in the linguistic system.

Regarding the role of cognates in processing OSV sentences, cognate verbs might be activated and selected for processing faster than non-cognate verbs, as a result of them having a higher current and resting activation level (cf. Section 1.1). Hence, I argue that cognates may lead to faster sentence processing and, consequently, to faster detection of a mismatch between an agent-first interpretation of the sentence and the non-linguistic context accompanying it (in other words, to faster realisation that the language has flexible word order). This is what I argue will trigger learning of agent-patient marking and, hence, of the target SOV and OSV structures. Thus, I propose that cognates may facilitate the process by which learning of agent-patient marking is prompted. In addition, cognates may facilitate the consolidation of the target SOV and OSV structures with postpositional agent-patient marking in learners' linguistic system, once they have been created and have received a low resting activation level. As mentioned, each time that a structure is processed its resting activation level increases and the extent of that increase depends on its current activation level. That is, the higher the current activation level of a structure is, the higher the resting level at which activation falls after processing (Sharwood Smith & Truscott, 2014, sec. 3.3.5). Since cognates have a higher current activation level than non-cognates, a higher activation will spread from cognate verbs than from non-cognate verbs to the structures containing them each time that they are processed. Consequently, the structures processed with cognates will have a higher current activation level than the structures processed with non-cognates and, hence, a higher resting activation level when processing terminates. The higher the resting activation level of a representation is, the more robustly established it is in the linguistic system (Sharwood Smith & Truscott, 2014, sec. 4.2). Thus, the two cross-linguistically dissimilar L2 structures should be more robustly established in learners' minds (i.e. should be better learnt) when processed in sentences with cognates as opposed to non-cognates.

Finally, I expect that the structure underlying SOV sentences and the one underlying OSV sentences will be comparably established in the linguistic system, irrespective of whether sentences are processed with a cognate or with a non-cognate verb. The two target structures, with verb-final word order and postpositional agent-patient marking, do not exist in learners' native language; they both have to be learnt from input. The subject-initial structure may be processed faster than the object-initial one, since the former can be correctly interpreted using the agent-first preference. However, the consolidation of the structures in the linguistic system (i.e. learning the verb-final word order and the agent-patient marking constituting them) should not be particularly harder for one structure compared to the other.

3. Experiment 4

3.1. Overview

The research question of Experiment 4 was "*Do cognates facilitate the initial acquisition of cross-linguistically dissimilar L2 structures?".* To address this question, I exposed two groups of Spanish natives without knowledge of Basque to the cross-linguistically dissimilar structures with a cognate verb or a non-cognate verb exemplified in Section 2.2. Then, I examined whether exposure to the structures with cognates, as opposed to non-cognates, facilitated the establishment of these structures in the linguistic system, as hypothesized in that section. I created two versions of a mini-language based on Basque. The two versions contained non-cognate nouns and either Spanish-Basque cognate verbs or non-cognate verbs, which were used to construct sentences exemplifying the target structures. As advanced, participants were two groups of Spanish natives. Those who learnt the structures with cognate verbs will be referred to as *cognate learners*; those who learnt them with non-cognate verbs will be referred to as *non-cognate learners.* The procedure of Experiment 4 was as follows. First, cognate and non-cognate learners learnt non-cognate nouns and either cognate or non-cognate verbs through picture-word association. Then, the two groups were exposed to the structures; they saw SOV and OSV sentences with postpositional agent-patient marking and with either a cognate or a non-cognate verb, each accompanied by a picture. After the exposure phase, cognate and non-cognate learners learnt novel non-cognate verbs through picture-word association. These verbs were used in the testing phase, where learning of the structures was tested in a sentence-picture congruency task and in a written production task. In the former, participants were presented with SOV and OSV sentence-picture pairs and had to indicate whether each sentence matched the picture accompanying it using their agent-patient marking knowledge and their vocabulary knowledge. In the production task, participants saw pictures of transitive actions and had to write a sentence that described each picture using the vocabulary and structures learnt. The experiment ended with a debriefing phase, where participants verbalized their syntax knowledge. My hypothesis was already advanced in Section 2.2 and can be summarised into two main claims. **First,** I hypothesized that the two L2 structures, with SOV and OSV word order and postpositional agent-patient marking, would be comparably established in learners' linguistic system. This would be the case when sentences were processed with cognate verbs and with non-cognate verbs. **Second,** I hypothesized that, overall, the structures would be more robustly established in cognate learners' linguistic system than in non-cognate learners' linguistic system. In familiar thinking terms, I hypothesized that cognates would facilitate the acquisition of the cross-linguistically dissimilar L2 structures and that this facilitation would be comparable for the two structures.

3.2. Participants

Sixty Spanish natives (51 female), divided into two groups of 30 cognate learners and 30 non-cognate learners, participated in the experiment. Their ages ranged from 18 to 31 (*M* = 20, *SD* = 2.97) and they were all students at the University of Barcelona (UB). All participants reported having no knowledge of Basque or other case-marking, verb-final languages in a linguistic background questionnaire. Most young Spanish-native adults living in Catalonia are bilingual with Catalan, which is the vehicular language in education (Law of linguistic policy 1/1998, article 20). Thus, participants were asked about their language history, proficiency and use in both languages using the same questionnaire as for Experiments 1-3, but adapted for Spanish-Catalan bilinguals (see Appendix C-1)[10]. Almost all cognate learners (93.33%) and non-cognate learners (90%) had at least some knowledge of Catalan. However, they all reported that they felt more comfortable using Spanish and that they were spoken to only in Spanish by at least one of their parents prior to starting school (0-3 years).

The questionnaire asked cognate and non-cognate learners about their frequency of use of Spanish and Catalan in three periods of their life: childhood (3-12 years), puberty (12-18 years) and adulthood (after 18 years) and in three environments: at school/university/work, at home

[10] Spanish and Catalan are typologically very close. Like Spanish, Catalan is an SVO, nominative-accusative language (Hualde, 1992), so I assumed that the fact that participants knew Catalan would not affect syntax learning. In addition, the cognate verbs used in the experiment were nearly full cognates in Basque, Spanish and Catalan, with Catalan verbs differing just slightly in the pronunciation of some vowels in respect to their Spanish translations.

and at other places. Responses were scored on a 7-point scale in which 1 corresponded to *Spanish only* and 7 to *Catalan only.* The mean scores for cognate and non-cognate learners in each life period and environment are reported in Appendix C-2. In brief, cognate and non-cognate learners were Spanish-dominant, for they currently used and had used Spanish the most throughout their lives (mean language use during childhood, cognate learners, 2.86 (*SD* = 1.63), non-cognate learners, 2.82 (*SD* = 1.63); puberty, cognate learners, 2.80 (*SD* = 1.49), non-cognate learners, 2.64 (*SD* = 1.54); adulthood, cognate learners, 2.68 (*SD* = 1.42), non-cognate learners, 2.54 (*SD* = 1.32)). The average frequency of use of Spanish was comparable between groups, as indicated by independent-samples t-tests (all $p \geq .40$, see Appendix C-2 for the details). Cognate and non-cognate learners were also asked about their self-rated proficiency speaking, listening, reading and writing in Spanish. Responses were scored on a 7-point scale in which 1 represented *very poor proficiency* and 7 represented *perfect proficiency.* The average scores for each skill reported by cognate and non-cognate learners can also be consulted in Appendix C-2. The two groups of participants considered they had a nearly perfect and comparable mastery of Spanish (mean proficiency collapsing all skills, cognate learners, 6.75 (*SD* = 0.49) vs. non-cognate learners, 6.80 (*SD* = 0.44), t (238) = -0.83, $p = .41$, $d = -0.11$). All participants reported having normal or corrected to normal vision and hearing. Before the experiment began, they read and signed an informed consent (Appendix C-3). Experiment 4 was part of the project “Cross-linguistic activation effects in bilingual language processing and learning” (PGC2018-097970-B-I00), funded by the Spanish Ministry of Science, Innovation and Universities and approved by the Committee of Ethics for research involving human beings of the University of the Basque Country (*Comité de Ética para las Investigaciones con Seres Humanos, CEISH,* Ref. M10_2019_167). Participants received 12€ for their participation.

3.3. Materials

I designed two versions of a mini-language based on Basque: a cognate version and a non-cognate version. The vocabulary of the cognate version consisted of five nouns denoting professions (*antzezle* “actor”*, epaile* “referee”*, sendagile* “doctor”*, margolari* “painter” and *gidari* “pilot”), four Spanish-Basque cognate verbs used in the exposure phase (*pintatu* “paint”*, salutatu* “greet”*, presentatu* “present” and *kastigatu* “punish”) and four non-cognate verbs used in the testing phase (*aukeratu* “choose”*, aztertu* “examine”*, zelatatu* “spy on” and *gainditu* “surpass”). The vocabulary of the non-cognate version was identical to that of the cognate version except that four non-cognate synonyms of the cognate verbs in the cognate version were used in the exposure phase (*margotu* “paint”*, agurtu* “greet”*, aurkeztu* “present” and *zigortu* “punish”). The vocabulary was selected using the dictionary *Euskaltzaindiaren Hiztegia*, created by the Royal Academy of the Basque Language (Euskaltzaindia, 2016). Each noun and verb was associated with a picture, bought from *123RF Image Bank* (https://www.123rf.com/). The full set of word-picture pairs can be consulted in Appendix C-4. These word-picture pairs were used in two vocabulary-learning phases: one preceding the exposure phase and one preceding the testing phase. They were also used to generate the sentence-picture pairs displayed in the exposure phase and in the testing phase (the exposure set and the testing set). The sentences of the exposure and testing sets were formed according to the two transitive structures differing in word order and agent-patient marking with Spanish described in Section 2.1. The pictures of transitive actions accompanying these sentences were created by manipulating the individual pictures of nouns and verbs bought from the image bank using *Pixelmator Pro* (Pixelmator Team, 2022, version 2.3.7). A male native speaker of Basque recorded the individual lexical items and the sentences of the exposure and the testing sets for use in the vocabulary-learning phases, the exposure phase and the testing phase. Recordings took place in a soundproof booth and using an Olympus voice recorder (Linear PCM Recorder LS-5 model, frequency sampling of 96kHz). Sentences were read at a normal pace and with natural intonation. The initial and final silences in each recording were cut using *Praat* (Boersma & Weenink, 2018, version 6.0.37). In what follows, I describe the exposure set and the testing set.

3.3.1. Exposure set

I generated 80 baseline sentences for the exposure set. These were all SOV sentences with postpositional agent-patient marking and with a cognate verb. Then, I generated three additional versions of each sentence manipulating word order (SOV vs. OSV) and the cognate status of the verb (cognate vs. non-cognate). The four versions of each sentence were paired with the same picture representing their meaning. The 160 sentence-picture pairs containing a cognate verb (80 SOV and their 80 OSV counterparts) were allocated to the cognate version of the language and the 160 sentence-picture pairs containing a non-cognate verb (80 SOV and their 80 OSV counterparts) were allocated to the non-cognate version of the language. The sentence-picture pairs for each language version were divided into two lists to prevent participants from hearing and seeing both an SOV sentence and its OSV counterpart. As a result, each participant was exposed to 80 sentence-picture pairs: 40 SOV and 40 OSV. Table 4.1 illustrates examples of equivalent SOV and OSV sentence-picture pairs for the cognate and the non-cognate language versions.

	Language version	**Sentence**	**Picture**
SOV	Cognate	Antzezle-ak gidari-a pintatu actor-A pilot-P paint "The actor is painting the pilot."	
	Non-cognate	Antzezle-ak gidari-a margotu actor-A pilot-P paint "The actor is painting the pilot."	

OSV	Cognate	Gidari-a antzezle-ak pintatu pilot-P actor-A paint "The actor is painting the pilot."	
	Non-cognate	Gidari-a antzezle-ak margotu pilot-P actor-A paint "The actor is painting the pilot."	

TABLE 4.1. Examples of SOV and OSV sentence-picture pairs for the cognate and the non-cognate version of the language in Experiment 4. The four sentences have the same meaning and, thus, they are paired with the same picture.

As mentioned in the previous section, the lexicon of the exposure set consisted of five non-cognate nouns and four verbs (cognate in the cognate language version, non-cognate in the non-cognate language version). The four cognate verbs and their Spanish counterparts had identical or nearly identical stems and different affixes (Spanish *–ar* vs. Basque *–tu*). To objectively measure the cognateness of verbs, I calculated the orthographic and phonological Levenshtein distance (LD) between the stems of the Basque verbs and their Spanish translations (orthographic LD, $M = 0.25$, $SD = 0.43$; phonological LD, $M = 0.5$, $SD = 0.5$). A LD of 0 indicates that two words are orthographically and/or phonologically identical. The higher the LD, the less similar two words are. The LD measures obtained confirmed that the verbs used were nearly identical Spanish-Basque cognates. Cognate and non-cognate verbs were matched in length (number of letters), as measured by an independent-samples t-test ($p > .05$). All nouns and verbs occurred with equal frequency in SOV and OSV sentences per list. Specifically, each noun occurred twice as the agent and twice as the patient of a given verb in SOV and OSV sentences per list. The combination of a given agent, patient and verb occurred only once in the same order in the exposure set. The complete exposure set is reported in Appendix C-5.

3.3.2. Testing set

3.3.2.1. Sentence-picture congruency task

The first task in the testing phase was a sentence-picture congruency task. The materials for this task were four different types of sentence-picture pairs shared for the cognate and the non-cognate language versions. Two of these sentence-picture pairs tested learning of the target syntactic structures. The other two tested vocabulary knowledge and were introduced to prevent participants from discovering the learning target during the congruency task. To create the four types of sentence-picture pairs, grammatical SOV and OSV sentences with postpositional agent-patient marking were associated with four different pictures:

1) An exact representation of the meaning of the sentence (*syntactically congruent sentence-picture pair – testing syntax learning*).

2) A picture that represented the reverse agent-patient relationship in the sentence (*syntactically incongruent sentence-picture pair – testing syntax learning*).
3) A picture that replaced the agent in the sentence by a different character (*semantically incongruent sentence-picture pair with agent violation – testing vocabulary learning*).
4) A picture that replaced the patient in the sentence by a different character (*semantically incongruent sentence-picture pair with patient violation – testing vocabulary learning*).

Table 4.2 provides examples of the four types of sentence-picture pairs constructed from an SOV sentence and its OSV counterpart. I first generated the maximum number of sentences that could be created with the vocabulary of the testing phase (the same five nouns as in the exposure set and four non-cognate verbs different from the ones in the exposure set) meeting the requisite that a given agent/subject, patient/object and verb occurred only once in the same order. These were 80 SOV sentences and 80 OSV sentences derived from the SOV ones (5 agents x 4 patients = 20 sentences x 4 verbs = 80 sentences). Each of these pairs of SOV and OSV sentences had to be associated with four pictures to yield the four types of sentence-picture pairs mentioned. This would result in eight different sentence-picture pairs, which would need to be distributed into eight lists of 80 sentence-picture pairs (40 SOV and 40 OSV: 10 syntactically congruent, 10 syntactically incongruent, 10 semantically incongruent with agent violation and 10 semantically incongruent with patient violation per word order).

	Pair type	**Sentence**	**Picture**
SOV	Syntactically congruent	Antzezle-ak gidari-a aukeratu actor-A pilot-P choose “The actor is choosing the pilot.”	
	Syntactically incongruent	Antzezle-ak gidari-a aukeratu actor-A pilot-P choose “The actor is choosing the pilot.”	
	Semantically incongruent with agent violation	Antzezle-ak gidari-a aukeratu actor-A pilot-P choose “The actor is choosing the pilot.”	
	Semantically incongruent with patient violation	Antzezle-ak gidari-a aukeratu actor-A pilot-P choose “The actor is choosing the pilot.”	

OSV	Syntactically congruent	Gidari-a antzezle-ak aukeratu pilot-P actor-A choose "The actor is choosing the pilot."	
	Syntactically incongruent	Gidari-a antzezle-ak aukeratu pilot-P actor-A choose "The actor is choosing the pilot."	
	Semantically incongruent with agent violation	Gidari-a antzezle-ak aukeratu pilot-P actor-A choose "The actor is choosing the pilot."	
	Semantically incongruent with patient violation	Gidari-a antzezle-ak aukeratu pilot-P actor-A choose "The actor is choosing the pilot."	

TABLE 4.2. Examples of the four types of sentence-picture pairs generated from an SOV sentence and its OSV equivalent in Experiment 4. The sentence-picture pairs were shared for the cognate and the non-cognate version of the language.

To this aim, I first associated each of the 80 SOV sentences and their 80 OSV counterparts with the pictures yielding syntactically congruent and incongruent sentence-picture pairs, for these could be derived straightforwardly from each sentence. Then, I divided the 80 syntactically congruent SOV sentence-picture pairs into eight groups of 10 and allocated each of these groups to a different list, creating, as required, eight lists. I did the same with the 80 syntactically incongruent SOV sentence-picture pairs and the 80 syntactically congruent and incongruent OSV sentence-picture pairs derived from the SOV pairs. A list never contained an SOV sentence and its OSV counterpart and never contained a syntactically congruent sentence-picture pair and its incongruent counterpart. Then, I needed to create four additional sentence-picture pairs from each SOV sentence (a semantically incongruent SOV sentence-picture pair with agent violation, a semantically incongruent SOV sentence-picture pair with patient violation and the two OSV counterparts). However, these sentence-picture pairs could not be derived straightforwardly from each sentence. For each picture, I needed to choose between four possible agent violations and four possible patient violations and this needed to be done in an organized manner. I did so in two steps. First, I took the 80 SOV *sentences* in the syntactically congruent sentence-picture pairs, divided them into eight groups of 10 and allocated each of these groups to a different list. The same went for the 80 SOV *sentences* in syntactically incongruent sentence-picture pairs and their OSV counterparts. In allocating these new groups of sentences to the eight existing lists, I made sure that a given SOV or OSV sentence would not appear twice in a list and that no list would contain an SOV sentence and its OSV counterpart. Second, I associated these SOV and OSV sentences with pictures to create the semantically incongruent sentence-picture pairs. The agents in the pictures constituting SOV or OSV sentence-picture pairs with agent violation were chosen from the agents in the SOV or OSV sentences in that condition and list, as long as doing so did not create an impossible sentence, i.e. a sentence having the same agent and patient. Likewise, the patients in the pictures constituting SOV or OSV sentence-picture pairs with patient violation were chosen from the patients in the SOV or OSV sentences in that condition and list, as long as doing so did not result in an impossible sentence. Pictures were shared for semantically incongruent SOV sentence-picture pairs with agent or patient violation and their OSV counterparts. The same picture only appeared once within condition per list.

In sum, the materials of the sentence-picture congruency task consisted of eight lists of 80 sentence-picture pairs (40 SOV and 40 OSV: 10 syntactically congruent, 10 syntactically incongruent, 10 semantically incongruent with agent violation and 10 semantically incongruent with patient violation per word order). As a reminder, the lexicon used to create the sentences consisted of five non-cognate nouns and four non-cognate verbs. Each noun occurred four times as the agent and four times as the patient of each verb per list: once in an SOV sentence-picture pair testing syntax learning, once in an OSV sentence-picture pair testing syntax learning, once in an SOV sentence-picture pair testing vocabulary learning and once in an OSV sentence-picture pair testing vocabulary learning. Whether these were syntactically congruent or incongruent sentence-picture pairs and semantically incongruent pairs with agent or patient violation was counterbalanced across lists. The materials of the sentence-picture congruency task can be consulted in Appendix C-5.

3.3.2.2. Production task

The second task in the testing phase was a picture-description task. I selected eight pictures from eight sentence-picture pairs used in the sentence-picture congruency task[11]. The same pictures were used for the cognate and the non-cognate language versions. Each of the four possible verbs appeared in two pictures and each of the five possible nouns fulfilled the role of agent and patient in at least one picture. Two characters never appeared together in more than one picture. The pictures can be consulted in Appendix C-5.

3.4. Procedure

Participants were told that they would learn some nouns and verbs in Basque and then do some sentence comprehension tasks. They were not informed that this was a syntax learning experiment. As advanced, the experiment had five phases: a first vocabulary-learning phase (Section 3.4.1), an exposure phase (Section 3.4.2), a second vocabulary-learning phase

[11] The pictures displayed in the sentence-picture congruency task represented the meaning of the maximum number of sentences that could be created with the vocabulary of the testing phase (5 nouns and 4 verbs). Because the vocabulary in the production task was the same as that in the sentence-picture congruency task, the pictures used in the former had to be a subset of those used in the latter.

(Section 3.4.3), a testing phase (Section 3.4.4) and a debriefing phase consisting of a verbal report (Section 3.4.5). Cognate and non-cognate learners additionally performed a reading span task (Section 3.4.6), which measures the ability to store and process information in short-term memory, to control for possible group differences in working memory capacity. The experiment was run on the E-Prime 3.0 software (Psychology Software Tools, Inc., 2016). Participants were tested one by one in a soundproof booth. All audio files were played through headphones. The experiment lasted for a maximum of an hour and a half. The instructions were in Spanish and can be found in Appendix C-6.

3.4.1. First vocabulary-learning phase

Cognate and non-cognate learners learnt the same five non-cognate Basque nouns (*antzezle* "actor", *epaile* "referee", *sendagile* "doctor", *margolari* "painter" and *gidari* "pilot"), but learnt four different verbs (Spanish-Basque cognates or their non-cognate equivalents, respectively, *pintatu/margotu* "paint", *salutatu/agurtu* "greet", *presentatu/aurkeztu* "present" and *kastigatu/zigortu* "punish"). Each trial began with a picture representing either a noun or a verb in the middle of a white screen. The Basque word associated with the picture was written below, together with its Spanish translation between brackets, and it was simultaneously played (Figure 4.1). Participants were instructed to repeat the word aloud and press the space bar, by which a new trial began. Each word-picture pair was repeated four times. The presentation of the pictures was pseudo-randomized for each participant, so that nouns were presented before verbs. Learning was assessed in a picture-word matching task and in a picture-naming task.

FIGURE 4.1. Example of a vocabulary learning trial in Experiment 4. The picture represents the Basque noun *gidari* ("pilot"), which was presented visually and aurally, as indicated by the speaker between brackets (not shown in the experiment). The Spanish translation "(piloto)" appeared below the noun.

3.4.1.1. Picture-word matching task

In each trial, participants were presented with a picture of a noun or a verb and a list of either the five nouns or the four verbs learnt written below (cognate verbs for cognate learners, non-cognate verbs for non-cognate learners). Participants were instructed to select the word that described the picture as quickly as possible using the numbers 1, 2, 3, 4 (or 5) on the keyboard (1 for the leftmost option, 2 for the option to its right, and so on). The number to be pressed to select each word appeared between brackets below each option. Following the selection of a word, feedback was provided for 700ms in the form of a green tick (correct answer) or a red cross (incorrect answer), after which a new trial began (Figure 4.2). The list of nouns or verbs appeared in a different order in each trial to prevent participants from predicting the correct response based on previous answers. Pictures of nouns appeared interspersed with pictures of verbs. If participants made a mistake in either a noun or a verb trial, they had to repeat the entire task again. The task was performed until 100% accuracy to make sure that cognate and non-cognate learners mastered the lexical items to the same extent. The presentation of the pictures was randomized so that, if participants had to perform the task more than once, they would not see the pictures in the same order.

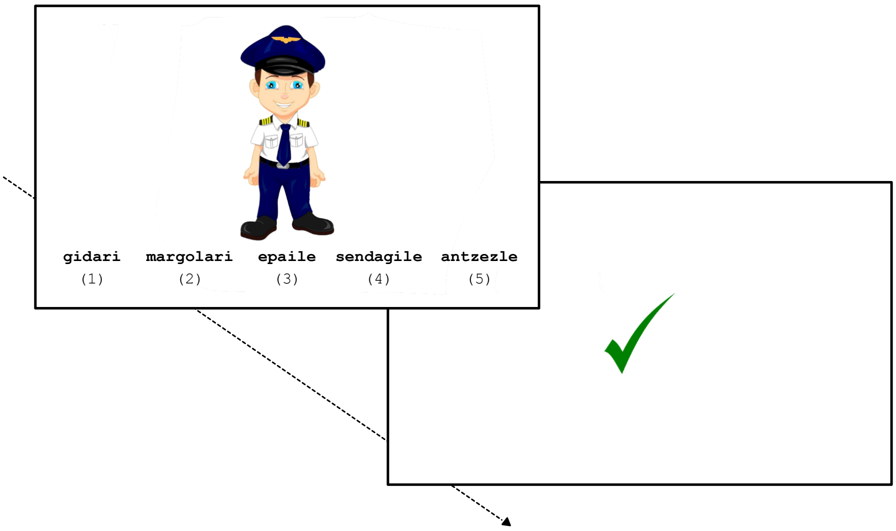


FIGURE 4.2. Example of a trial in the picture-word matching task in Experiment 4. The picture represents the Basque noun *gidari* ("pilot"). Below are written the five nouns of the mini-language and the numbers that had to be pressed to select each word. After selecting a noun, feedback was shown for 700ms.

3.4.1.2. Picture-naming task

Cognate and non-cognate learners saw the pictures of nouns and verbs one at a time and were asked to name them as quickly and accurately as possible. After naming a picture, participants pressed the space bar to see and hear the correct response. Each picture had to be named within 10 seconds. If no answer was given in this time, the correct response automatically appeared written below the picture. The presentation of the pictures was pseudo-randomized, so that nouns had to be named before verbs. The computer recorded participants' responses. In addition, I monitored their utterances to check in real time whether they could name the pictures appropriately. If participants named all pictures correctly, they moved on to the exposure phase. If they had difficulty naming some pictures, I reviewed the vocabulary with them and they performed the picture-naming task again. This procedure was repeated until cognate and non-cognate learners named all pictures correctly.

3.4.2. Exposure phase

Cognate and non-cognate learners were informed that they would see pictures involving some of the nouns and verbs learnt, each visually and aurally accompanied by a sentence describing it. They were told that they simply had to look at each picture, and listen to and read the sentence accompanying it. Cognate learners were exposed to SOV and OSV sentences with postpositional agent-patient marking and with cognate verbs; non-cognate learners were exposed to the same sentences but with non-cognate verbs. Participants were not informed that they would be exposed to two different structures. Each picture appeared in the middle of a white screen. The sentence that described it was written below and it was simultaneously played (Figure 4.3). Each sentence-picture pair remained on the screen for 500ms after the audio of the sentence ended. Then, a new pair was automatically presented. The order of the agent and the patient in the pictures was counterbalanced so that, in half of the pictures, the agent appeared on the right and in the other half, on the left. The order of appearance of sentence-picture pairs was randomized for each participant.

FIGURE 4.3. Example of an exposure trial for the non-cognate language version in Experiment 4. Participants saw a picture while reading and listening to a sentence describing it. The speaker between brackets (not shown in the experiment) indicates that the sentence was aurally presented. The figure displays an SOV sentence-picture pair representing "The actor is painting the pilot" with the non-cognate verb *margotu* ("paint").

3.4.3. Second vocabulary-learning phase

Cognate and non-cognate learners learnt the same four novel non-cognate verbs (*aukeratu* "choose", *aztertu* "examine", *zelatatu* "spy on" and *gainditu* "surpass"). Learning new non-cognate verbs was necessary to be able to later test cognate and non-cognate learners with the same materials. This would allow judging the extent to which being exposed to the target structures in sentences with cognates as opposed to non-cognates had influenced syntax learning. The procedure by which the verbs were learnt and tested was identical to the one in the first vocabulary-learning phase.

3.4.4. Testing phase

3.4.4.1. Sentence-picture congruency task

Participants were told that they would see sentence-picture pairs including the verbs they had just learnt, but that apart from this lexical change, the sentences and pictures would be like the ones in the exposure phase. They were informed that the sentence-picture pairs would appear one at a time and that they would need to decide, as quickly as possible, whether each sentence was a correct description of the picture or not. As a reminder, participants had to judge four types of SOV and OSV sentence-picture pairs: (i) syntactically congruent, (ii) syntactically incongruent, (iii) semantically incongruent with agent violation and (iv) semantically incongruent with patient violation. The first two pairs tested syntax learning. The last two pairs tested vocabulary learning and were included to avoid that participants discovered the target structures while conducting the congruency task. Each trial started with a picture in the middle of a white screen and a sentence written below, which was simultaneously played. The order of the agent and the patient in the pictures was counterbalanced: in half of the pictures, the agent was on the right and in the other half, it was on the left. When the audio of the sentence stopped playing, the options *Correcto* ("Correct") and *Incorrecto* ("Incorrect") appeared below the sentence-picture pair, which remained above the two (Figure 4.4). Participants had to press the key "A" to indicate that the sentence was a correct description of the picture and the key "L" to indicate that the sentence was an incorrect description of the picture. Participants had 5 seconds to respond. If after this time no choice was made, a message indicating that no response was detected was displayed and a new trial began. No feedback was provided regarding the accuracy of responses. The presentation of the sentence-picture pairs was randomized for each participant. Prior to the task, participants conducted a short practice to familiarize themselves with this task. This included four sentence-picture pairs (two SOV and two OSV) randomly chosen from the testing set: one syntactically congruent, one syntactically incongruent, one semantically incongruent with agent violation and one semantically incongruent with patient violation[12].

[12] Remember that the maximum number of sentences that could be created with the vocabulary of the testing phase were used as experimental sentence-picture pairs. Hence, in the practice session some pairs randomly chosen from the experimental ones had to be used.

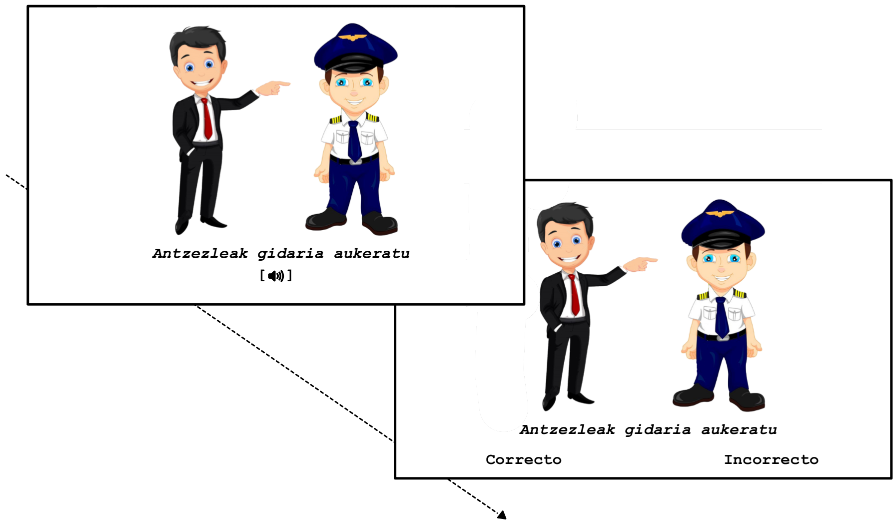


FIGURE 4.4. Example of a trial in the sentence-picture congruency task in Experiment 4. First, participants saw a picture while reading and listening to a sentence. The speaker between brackets (not shown in the experiment) indicates that the sentence was aurally presented. After the audio of the sentence ended, the options *Correcto* ("Correct") and *Incorrecto* ("Incorrect") appeared below the sentence-picture pair and participants judged the congruency between the two. In this figure, the sentence means "The actor is choosing the pilot" and it is a correct description of the picture.

3.4.4.2. Production task

Participants were told that they would see some pictures similar to the ones in the previous task and that they would have to write a sentence that described each picture in a text box. To help them remember how the nouns and verbs were written, the five possible (bare) nouns and the four possible verbs were listed to the right of the picture (Figure 4.5). Cognate and non-cognate learners saw the same eight pictures in a randomized order. The appearance of the agent and the patient on the right or the left of the picture was counterbalanced. There was no time limit for participants to type their answers. When they finished writing a sentence, or if they did not know how to describe a picture, they pressed ENTER and a new trial began.

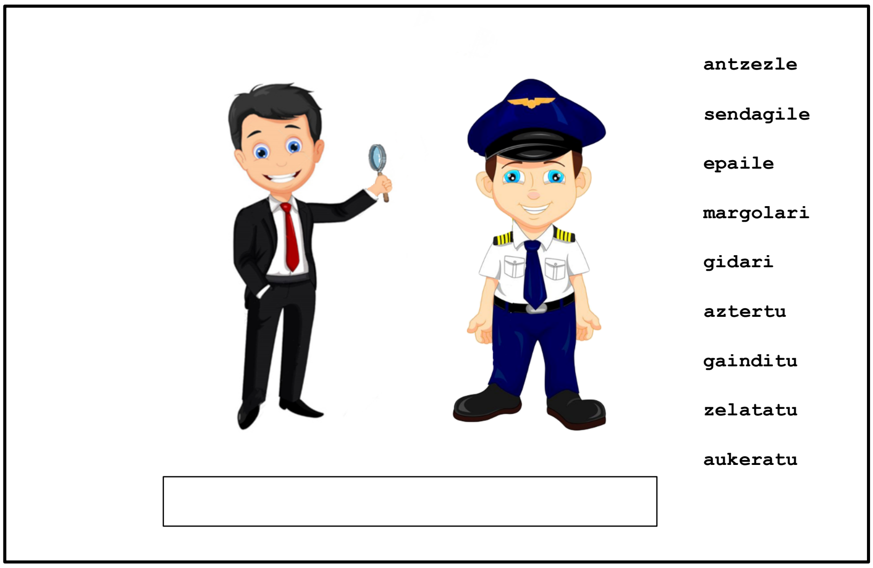


FIGURE 4.5. Example of a trial in the production task in Experiment 4. Participants saw a picture and had to describe it choosing the appropriate nouns and verb from the list and using one of the structures learnt. In this figure, possible picture descriptions were *Antzezleak gidaria aztertu* (SOV) or *Gidaria antzezleak aztertu* (OSV) ("The actor is examining the pilot").

3.4.5. Debriefing phase

Participants were encouraged to verbalize any knowledge about the agent-patient marking in the mini-language. In addition, they were asked about the part of the experiment in which they became aware of this marking to make sure that, at least for those participants who developed conscious syntax knowledge, learning was the result of the exposure phase. Finally, participants were asked about any strategies used to conduct the sentence-picture congruency task (e.g. intuition, syntax knowledge or other). Specifically, they were asked:

1. In all the sentences that you have heard, there were two nouns (*antzezle, gidari, sendagile, margolari* or *epaile*). Did you notice that when these nouns appeared in a sentence their original form changed?
2. Could you say how did it change?
3. Could you say why?
4. In which part of the experiment did you notice the change in form?
5. Did you follow any strategy to perform the test?

The questions appeared on the screen accompanied by a text box in which participants typed their answers. In addition, I read the questions with them and made sure that they understood that the "test" in the fifth question referred to the sentence-picture congruency task. If they could not think of any strategies used to perform this test, I asked them whether they had followed their intuition or a conscious criterion to judge the sentence-picture pairs as correct or incorrect. On the one hand, learning occurred under incidental conditions (i.e. participants were exposed to the language without being informed that they had to learn two structures or that they would be tested on their syntax knowledge afterwards, Rebuschat, 2013). These learning conditions should favour the development of unconscious syntax knowledge, so participants were not expected to provide metalinguistic descriptions of agent-patient marking. On the other hand, even if during the experiment learners were not informed that they had to learn some structures nor were they encouraged to look for patterns in the input, they could have done so for a number of reasons. In the exposure phase, the instructions were simply to look at each picture and listen to and read the accompanying sentence. Yet, it is possible that by focusing their attention on the sentence-picture pairs, participants consciously or unconsciously focused on the form of the sentences too. This is even more likely if we consider that L2 learners have already acquired the native language and possess metalinguistic awareness, "the ability to focus attention on language as an object in and of itself, to reflect upon language, and to evaluate it" (Thomas, 1988, p. 531). Considering all this, it is also possible that participants developed conscious knowledge of the agent-patient marking in the language. In this case, an expected description of this marking was:

- When the nouns appeared in a sentence, they had the suffixes *–ak* or *–a* attached. When a noun ended in *–ak*, it was the agent/subject of the sentence. When a noun ended in *–a*, it was the object/patient of the sentence.

3.4.6. Reading span task

To confirm that cognate and non-cognate learners had comparable working memory capacities, both groups performed the Spanish version of Unsworth et al.'s (2005) reading span task, which requires participants to remember individual letters while performing plausibility judgements (Figure 4.6). The task had four parts. In the first part, the two groups were trained on letter recall using a letter span. Participants saw 10 capital letters one by one presented in two sets of two letters and two sets of three letters. They were instructed to remember them in their order of presentation. Each letter appeared on the screen for 800ms. At the end of each set, a 4x3 matrix of letters appeared and participants had to indicate the order of appearance of the letters by clicking on the box next to each letter. Feedback was provided on the number of letters correctly recalled. In the second part, participants practised making plausibility judgements. They read aloud 15 sentences (7 plausible and 8 implausible) one at a time. After the presentation of each sentence, they clicked on the screen and the sentence was replaced by the message "This sentence makes sense" accompanied by the options *True* and *False*. Participants made their judgement by clicking on one of the options. Feedback was provided on the correctness of their judgements. In the third part, participants practised performing the reading span task, i.e. recalling letters while reading sentences and judging their plausibility. Each trial started with a sentence to read, after which participants conducted their plausibility judgement as described above. Following each judgement, a to-be-recalled letter appeared on the screen for 800ms, after which it was replaced by the following sentence. Participants had to remember six letters presented in three sets of two letters each. When a set of letters had been presented, the 4x3 matrix of letters was displayed and participants had to recall the order of appearance of the letters. The last part of the reading span task was the test itself. This was identical to the practice conducted in the third part, but this time participants had to recall five blocks of three sets of letters ranging in size from three to seven. The order of appearance of the sets was randomized. Since it was crucial that participants attempted to both perform plausibility judgements and recall the letters (instead of focusing just on remembering the letters), they were asked to keep a minimum accuracy of 85% in their judgements. The percentage of sentences correctly judged appeared in red on the upper right-hand corner of the screen. The task took about 20 minutes to complete.

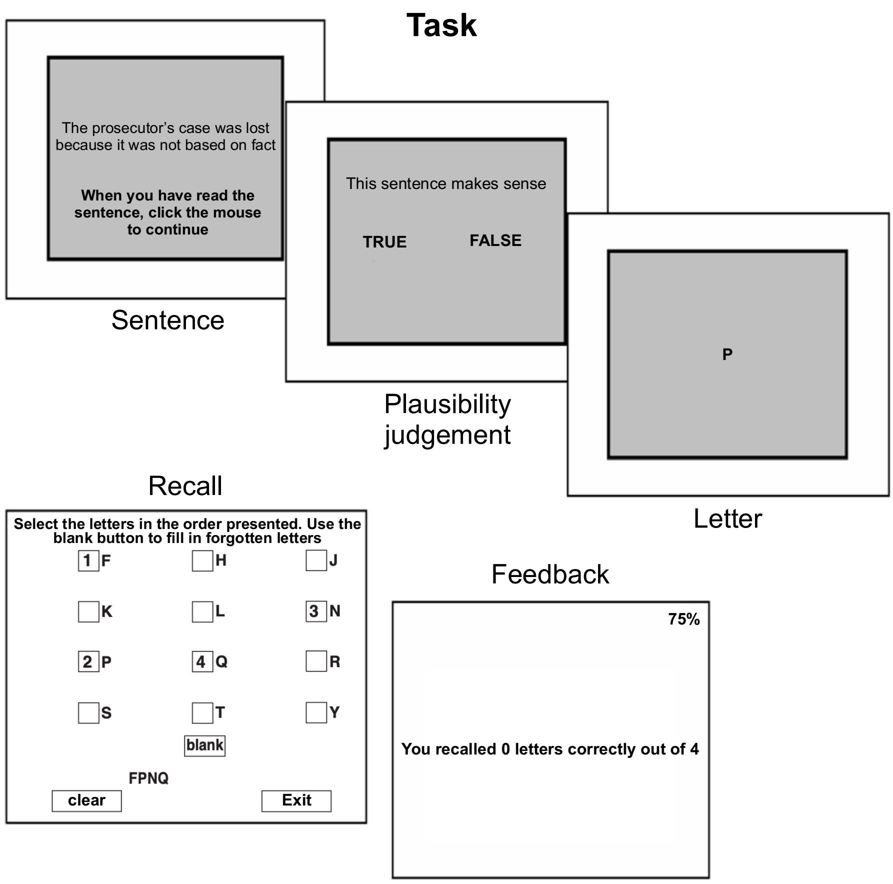


FIGURE 4.6. Experimental procedure in the reading span task (adapted from Unsworth et al., 2005).

3.5. Predictions

In Section 3.1, I presented my hypothesis about how exposure to the two cross-linguistically dissimilar L2 structures with cognate verbs vs. non-cognate verbs would influence the initial establishment of these structures in learners' linguistic system. Experiment 4 was designed to test this hypothesis, which, in turn, addressed the research question of the study (*"Do cognates facilitate the initial acquisition of cross-linguistically dissimilar L2 structures?"*). In this section, I will review the hypothesis and present its predictions for Experiment 4. In addition, during the experiment cognate and non-cognate learners learnt non-cognate nouns, cognate or non-cognate verbs (first vocabulary-learning phase) and novel non-cognate verbs (second vocabulary-learning phase). Thus, some predictions could also be outlined for this vocabulary learning. In what follows, I will first present my predictions for vocabulary learning and then my predictions for syntax learning.

Predictions for vocabulary learning

In the first and second vocabulary-learning phases, cognate and non-cognate learners learnt the vocabulary to the same extent (until reaching 100% accuracy in the picture-word matching task and in the picture-naming task). In the first vocabulary-learning phase, the two groups learnt the same non-cognate nouns. Thus, I predicted that cognate and non-cognate learners would correctly match all nouns to pictures (picture-word matching task) and would correctly name all pictures of nouns (picture-naming task) in a comparable number of attempts. Learning of nouns was also assessed in the testing phase, particularly in the sentence-picture congruency task, when learners were asked to judge the congruency of semantically incongruent sentence-picture pairs in which the characters acting as agent or patient in the sentence did not match the ones in the picture. Even if these pairs were introduced to prevent participants from discovering the target structures while judging syntactically congruent and incongruent sentence-picture pairs, they could further corroborate that learning of nouns was comparable for cognate and non-cognate learners. If this was the case, I predicted that accuracy in semantically incongruent sentence-picture pairs would be comparable for the two groups of learners. I created semantically incongruent sentence-picture pairs with either an agent or a patient violation so that, in order to perform their congruency judgements, learners had to pay attention to both the agent and the patient in each sentence, just as they had to do in syntactically congruent and incongruent sentence-picture pairs. I did not predict a statistically significant difference in accuracy when judging sentence-picture pairs with agent vs. patient violation in neither of the two groups of learners, since doing this only required accessing semantic knowledge, which would be comparable between groups.

Going back to the first vocabulary-learning phase, apart from learning nouns, cognate learners learnt cognate verbs and non-cognate learners learnt non-cognate verbs. Due to the cross-linguistic similarity of the cognates studied in Basque and in Spanish (participants' L1), I predicted that cognates would be easier to learn than non-cognates. That is, I predicted that cognate learners would correctly match all verbs to pictures (picture-word matching task) and/or would correctly name all pictures of verbs (picture-naming task) in significantly fewer attempts than non-cognate learners would. Finally, in the second vocabulary-learning phase, cognate and non-cognate learners learnt the same four non-cognate verbs. Like in the case of nouns in the first vocabulary-learning phase, I predicted that the two groups would not statistically differ in the number of attempts necessary to reach 100% accuracy in the picture-word matching task and in the picture-naming task.

Predictions for syntax learning

My hypothesis about the facilitative role of cognates in the initial acquisition of the cross-linguistically dissimilar L2 structures included two main claims. **First,** I claimed that the two structures, with SOV and OSV word order and postpositional agent-patient marking, would be comparably established in the linguistic system. This would be the case when the structures were processed with cognate verbs and with non-cognate verbs. **Second,** I claimed that, overall, the structures would be more robustly established in cognate learners' linguistic system than in non-cognate learners' linguistic system. A prerequisite for the hypothesis to be met is that the two target structures were established in cognate and non-cognate learners' linguistic system. This could be seen in different ways in the sentence-picture congruency task and in the production task.

As mentioned in Chapter 2, adults, who know how their native language grammar works, can tell apart grammatical sentences (acceptable in the language) from ungrammatical ones (unacceptable in the language) (Chomsky, 1965). A similar reasoning could be made for L2 learners. In the congruency task, sentences in syntactically congruent sentence-picture pairs were acceptable in the L2, for agent-patient marks were used correctly, i.e. the nouns marked as the agent and the patient in the sentence corresponded to the agent and the patient in the picture. By contrast, sentences in syntactically incongruent sentence-picture pairs could be considered unacceptable in the L2, for agent-patient marks were used incorrectly, i.e. the nouns marked as the agent and the patient in the sentence had the opposite conceptual roles in the picture. If the SOV and the OSV structures with postpositional agent-patient marking were part of learners' linguistic system, I predicted that learners would be able to tell apart syntactically congruent SOV and OSV sentence-picture pairs from syntactically incongruent SOV and OSV sentence-picture pairs. In other words, cognate and non-cognate learners would obtain above-chance accuracy when judging these four types of sentence-picture pairs. Turning to the production task, if the two structures were part of learners' linguistic system, I made three predictions. First, I predicted that cognate and non-cognate learners would write SOV and OSV sentences. Second, I predicted that they would use agent-patient marking in significantly more than 50% of SOV and OSV sentences. Finally, I predicted that learners would use this marking correctly in significantly more than 50% of SOV and OSV sentences with agent-patient marks.

If these predictions were met, I will then test the two claims of my hypothesis. As mentioned, **the first claim of the hypothesis** was that the SOV and the OSV structures would be comparably established in the linguistic system and that this would be the case for cognate and non-cognate learners. **If this first claim was correct,** I predicted that, in the congruency task, accuracy when judging syntactically congruent and incongruent SOV vs. OSV sentence-picture pairs would not vary as a function of group of learners. In addition, overall, cognate and non-cognate learners would not be significantly more accurate when judging one or the other type of sentence-picture pairs. **The second claim of the hypothesis** was that, overall, the structures would be more robustly established in cognate learners' linguistic system than in non-cognate learners' linguistic system. **If this second claim was correct,** I predicted that accuracy when judging sentence-picture pairs would be significantly higher for cognate learners than for non-cognate learners. Additionally or alternatively, similar results would be observed in the production task. Specifically, **if the first claim of the hypothesis was true,** the accuracy with which learners wrote SOV vs. OSV sentences with agent-patient marking would not significantly change as a function of group of learners and, overall, cognate and non-cognate learners would not be significantly more accurate when producing one or the other type of sentences. On the other hand, **if the second claim of the hypothesis was true**, cognate learners would be significantly more accurate in their verb-final picture descriptions with agent-patient marking than non-cognate learners would.

3.6. Coding and data analysis

Experiment 4 was analysed using the programming environment R (R Core Team, 2022, version 4.2.2). The function and package used to calculate each statistical test and effect size measure is only stated the first time that a test or effect size is reported.

3.6.1. First vocabulary-learning phase

3.6.1.1. Picture-word matching task

Accuracy was measured after the selection of a word in each trial. Responses were coded as binary (1 = correct picture-word matching, 0 = incorrect picture-word matching). Additionally, I coded the number of attempts at picture-noun matching and picture-verb matching per participant as positive integers. Shapiro-Wilk tests calculated with the function *shapiro.test* from the *stats* package (R Core Team, 2022) indicated that the number of attempts at picture-noun matching were not normally distributed neither for cognate learners ($W = 0.62$, $p < .001$) nor for non-cognate learners ($W = 0.74$, $p < .001$), as it is usually the case with discrete variables. Likewise, the number of attempts at picture-verb matching were not normally distributed nor for cognate learners ($W = 0.69$, $p < .001$) nor for non-cognate learners ($W = 0.78$, $p < .001$). Because this non-normality could decrease the reliability of a parametric test, I compared cognate and non-cognate learners' number of attempts at picture-noun matching, on the one hand, and at picture-verb matching, on the other hand, using the non-parametric version of the independent-samples t-test, the Wilcoxon rank-sum test. This test assesses whether the two populations compared (in this case, the samples of attempts at picture-word matching for cognate and non-cognate learners) have equal distributions with the same median. In other words, whether one distribution has significantly larger values than the other does (Hogg et al., 2015). Wilcoxon rank-sum tests were computed using the function *wilcox.test* from the *stats* package. Continuity correction was used to reduce the risk of Type I error (Levshina, 2015). I calculated the standardised measure of effect size *r* for this test in this and subsequent tasks, using the function *wilcox_effsize* from the *rstatix* package (Kassambara, 2021). An *r* of 0.1-0.3, 0.3-0.5 and ≥ 0.5 was considered small, medium and large, respectively.

3.6.1.2. Picture-naming task

As in the previous task, I coded as positive integers the number of attempts necessary for each participant to name correctly all pictures of nouns and all pictures of verbs. Cognate and non-cognate learners' number of attempts to name noun pictures were not normally distributed (cognate learners, $W = 0.85$, $p < .001$; non-cognate learners, $W = 0.81$, $p < .001$). The two groups' number of attempts to name verb pictures were also non-normally distributed (cognate learners, $W = 0.45$, $p < .001$; non-cognate learners $W = 0.81$, $p < .001$). I used the non-parametric Wilcoxon rank-sum test with continuity correction to compare cognate and non-cognate learners' number of attempts at noun picture naming, on the one hand, and at verb picture naming, on the other hand.

3.6.2. Second vocabulary-learning phase

3.6.2.1. Picture-word matching task

Accuracy was measured following the selection of a word in each trial. Responses were binary-coded (1 = correct picture-word matching, 0 = incorrect picture-word matching). The number of attempts at picture-word matching per participant were coded as positive integers. The mean and median number of attempts was the same for cognate and non-cognate learners. A Wilcoxon rank-sum test with continuity correction compared the distribution of the number of attempts of the two groups of learners to see whether these additionally had similar shapes, i.e. whether one of the two distributions had significantly larger values than the other one.

3.6.2.2. Picture-naming task

I coded the number of attempts at picture naming per participant as positive integers. According to Shapiro-Wilk tests, the number of attempts needed to name pictures correctly were non-normally distributed both for cognate learners ($W = 0.53$, $p < .001$) and for non-cognate learners ($W = 0.35$, $p < .001$). I used the non-parametric Wilcoxon rank-sum test with continuity correction to compare the two groups' number of attempts at the task.

3.6.3. Testing phase

3.6.3.1. Sentence-picture congruency task

In this section, I describe how I coded and analysed the results of the sentence-picture congruency task. This includes the strategies that participants followed to perform the task, even if this information was reported in the debriefing phase.

<u>Performance on semantically incongruent sentence-picture pairs</u>

I measured accuracy after judging each sentence-picture pair. This accuracy was coded as binary (1 = correct sentence-picture congruency judgement, 0 = incorrect sentence-picture congruency judgement). Trials in which participants did not provide a response were eliminated, since E-prime coded accuracy in these trials as 0 even though participants made no judgement. Overall, I eliminated 4.17% (50/1200) of all sentence-picture pairs seen by cognate learners and 4% (48/1200) of all pairs seen by non-cognate learners. As a reminder, all semantically incongruent sentence-picture pairs contained a mismatch between either the agent or the patient in the sentence and the ones in the picture. I analysed whether accuracy in these two types of sentence-picture pairs was comparable for cognate and non-cognate learners, which would indicate comparable knowledge of nouns between groups. Specifically, a generalized linear mixed effects model fitted with the function *glmer* from the *lme4* package (D. Bates, Mächler, et al., 2015) tested for the interaction between the effect of Group of learners (Cognate vs. Non-cognate) and Type of sentence-picture pair (with Agent violation vs. with Patient violation) on accuracy. The model had random intercepts by participant and by item, and a random slope of Type of sentence-picture pair by participant. I used deviation coding for the variables Group of learners (*Cognate* coded as 0.5 and *Non-cognate*, as -0.5) and Type of sentence-picture pair (*with Agent violation* coded as 0.5 and *with Patient violation*, as -0.5)[13].

Performance on syntactically congruent and incongruent sentence-picture pairs

I measured accuracy after judging each sentence-picture pair. This was binary-coded (1 = correct sentence-picture congruency judgement, 0 = incorrect sentence-picture congruency judgement). Trials in which participants provided no response were eliminated. These were 4.08% (49/1200) of all sentence-picture pairs seen by cognate learners and 4.5% (54/1200) of all pairs seen by non-cognate learners. For cognate learners, this corresponded to 4.67% (14/300) of all syntactically congruent SOV sentence-picture pairs, 3% (9/300) of all syntactically congruent OSV sentence-picture pairs, 6% (18/300) of all syntactically incongruent SOV sentence-picture pairs and 2.67% (8/300) of all syntactically incongruent OSV sentence-picture pairs. For non-cognate learners, this corresponded to 3.67% (11/300) of all syntactically congruent SOV sentence-picture pairs, 6.67% (20/300) of all syntactically congruent OSV sentence-picture pairs, 5% (15/300) of all syntactically incongruent SOV sentence-picture pairs and 2.67% (8/300) of all syntactically incongruent OSV sentence-picture pairs.

The structures as part of the linguistic system

As a first step, I assessed whether the two target structures, with verb-final word order and postpositional agent-patient marking, were part of learners' linguistic system. I descriptively and visually examined cognate and non-cognate learners' mean accuracy percentages when judging syntactically congruent SOV sentence-picture pairs, syntactically congruent OSV sentence-picture pairs, syntactically incongruent SOV sentence-picture pairs and syntactically incongruent OSV sentence-picture pairs. All graphs in this chapter were generated with the function *ggplot* from the *ggplot2* package (Wickham, 2016). One-sample t-tests fitted with the function *t.test* from the *stats* package compared each percentage against chance (50%). *Cohen's d* was calculated as a standardised measure of effect size for these t-tests (and for all

[13] As mentioned in previous chapters, this coding scheme eases the interpretation of the effects of the independent variables on the dependent variable as main effects in the presence of an interaction (Sonderegger et al., 2018).

t-tests in this chapter) using *cohens_d* from the *rstatix* package. A *d* of 0.2, 0.5 and 0.8 was considered small, medium and large, respectively (Cohen, 1988). The descriptive and visual analysis of mean accuracy percentages suggested that cognate and non-cognate learners were biased towards judging as "correct" both syntactically congruent and incongruent sentence-picture pairs. To corroborate this, I coded responses in the test as *Hits* (congruent sentence-picture pair judged as "correct"), *False alarms* (incongruent sentence-picture pair judged as "correct"), *Misses* (congruent sentence-picture pair judged as "incorrect") or *Correct rejections* (incongruent sentence-picture pair judged as "incorrect"). Then, I calculated the index of response bias *c* from the Signal Detection Theory for (i) syntactically congruent and incongruent SOV sentence-picture pairs and (ii) syntactically congruent and incongruent OSV sentence-picture pairs for both groups of participants. These were calculated for each participant individually with the function *dprime* from the *psycho* package (Makowski, 2018). Then, mean indices were compared against zero using one-sample t-tests. The analysis confirmed that cognate and non-cognate learners were biased towards accepting SOV and OSV sentence-picture pairs, irrespective of their congruency[14].

To examine whether this bias affected cognate and non-cognate learners equally, a generalized linear mixed effects model looked into the interaction between the effect of Group of learners (Cognate vs. Non-cognate) and Word order of the sentence-picture pair (SOV vs. OSV) on accuracy. I pooled together accuracy for (i) syntactically congruent and incongruent SOV sentence-picture pairs and (ii) syntactically congruent and incongruent OSV sentence-picture pairs. The model that better fitted the data included random intercepts by participant and by item and a by-participant random slope of Word order of the sentence-picture pair, as indicated by nested model comparisons[15]. Deviation coding was used for the variables Group of learners and Word order of the sentence-picture pair. The categories *Cognate* and *SOV* were assigned the value 0.5 and the categories *Non-cognate* and *OSV* were assigned the value -0.5.

<u>Strategies used to perform the sentence-picture congruency task</u>

Finally, to have a better understanding of participants' performance, I descriptively analysed the strategies that cognate and non-cognate learners reported following to perform the sentence-picture congruency task (to be consulted in Appendix C-7). As will be discussed in Section 3.7.3, the analysis revealed that the aforementioned response bias could partially reflect a conscious strategy to judge sentence-picture pairs based only on the congruency between the nouns in the sentence and the picture, irrespective of agent-patient marking. In this case, sensitivity to the difference between syntactically congruent and incongruent

[14] Remember that the index of bias *c* is defined as the negative value of half of the sum of the z-transforms of the hit rate (in this case, the probability of judging a congruent sentence-picture pair as "correct") and the false alarm rate (in this case, the probability of misjudging an incongruent sentence-picture pair as "correct"). A *c* value of 0 reflects no bias to accept or reject the stimuli. When *c* is significantly higher than 0, it indicates a bias towards rejecting most stimuli, whereas when *c* is significantly lower than 0, it reflects a bias towards accepting them.

[15] Comparison of the models with and without the by-participant random slope: $X^2(2) = 156.46$, $p < .001$.

sentence-picture pairs would be impossible to assess in the congruency task, not even with a *d'* analysis, used to isolate sensitivity from *unconscious* response bias. Because the results of the task were not reliable, it was analysed no further.

3.6.3.2. Production task

The structures as part of the linguistic system

Like in the sentence-picture congruency task, I first examined whether the two target structures were established in cognate and non-cognate learners' linguistic system. In the production task, participants had to write sentences to describe pictures of transitive actions using the vocabulary learnt. I coded cognate and non-cognate learners' picture descriptions as *SOV, SVO, OSV, OVS* or *Not available (NA)* looking only at the order of the nouns and the verb in the sentence (irrespective of whether agent-patient marking was used or not). A sentence was considered subject-initial if the noun of the subject/agent in the picture was placed in sentence-initial position. A sentence was considered object-initial if the noun of the object/patient in the picture was placed in sentence-initial position. Two trials (one from a cognate learner, one from a non-cognate learner) were coded as NA and were removed from the analysis. These corresponded to a trial in which no picture description was written and a trial in which the same noun was used as the subject and as the object of the sentence. Next, those trials in which participants wrote verb-medial picture descriptions were eliminated. Only non-cognate learners wrote this type of sentences (12.97%, 31/239 SVO sentences; 2.93%, 7/239 OVS sentences). With the remaining data, I calculated the proportion of SOV and OSV picture descriptions written by cognate and non-cognate learners as evidence that both groups had learnt that the two word orders were possible in the mini-language.

Next, I coded agent-patient marking use in these picture descriptions as a binary variable (1 = sentence with agent-patient marking, 0 = sentence without agent-patient marking). A sentence was considered to contain agent-patient marking if: (i) participants had attached the agent and the patient mark to its two nouns, (ii) participants had only attached one of the two marks to one of its nouns or (ii) participants had attached the same mark to its two nouns. I calculated the proportion of SOV and OSV picture descriptions with and without agent-patient marking written by cognate and non-cognate learners. One-sample t-tests assessed whether the two groups used agent-patient marks in significantly more than 50% of SOV and OSV sentences, which would indicate that they had learnt that there was agent-patient marking in the language. Finally, I took the subset of picture descriptions with agent-patient marking and I coded its correct/incorrect use as a binary variable (1 = correct agent-patient marking, 0 = incorrect agent-patient marking). A sentence was considered to have correct agent-patient marking if the agent mark was attached to the noun acting as agent in the picture and/or if the patient mark was attached to the noun acting as patient in the picture. One-sample t-tests assessed whether the two groups of learners used agent-patient marks as required in significantly more than 50% of these SOV and OSV sentences, which would show that learners knew how agent-patient marking was used in the language. As will be detailed in Section 3.7.3, the analysis revealed that the two structures, with verb-final word order and postpositional agent-patient marking, were established in cognate and non-cognate learners' linguistic system. In this light, I turned to examining whether being exposed to the structures in sentences with cognate verbs as opposed to non-cognate verbs facilitated their acquisition, as predicted by my hypothesis.

Comparing the establishment of the structures in cognate and non-cognate learners' linguistic system

I assessed whether cognate and non-cognate learners' accuracy when writing SOV and OSV picture descriptions with agent-patient marking significantly differed. A generalized linear mixed effects model tested for an interaction between the effect of Group of learners (Cognate vs. Non-cognate) and Word order of the picture description (SOV vs. OSV) on accuracy. The model that converged had random intercepts by participant and by item. I used deviation coding for the variable Group of learners (*Cognate* coded as 0.5; *Non-cognate,* as -0.5) and the variable Word order of the picture description (*SOV* coded as 0.5; *OSV*, as -0.5).

3.6.4. Debriefing phase

I transcribed cognate and non-cognate learners' responses in the verbal report. Learners were first asked whether they had noticed that when nouns appeared in a sentence their original form changed. If their response was affirmative, they were asked what this change was and why it occurred. I calculated the percentage of cognate and non-cognate learners who (i) reported noticing that nouns in a sentence were different from their citation form, (ii) reported that (one or several) specific marks were affixed to nouns and (iii) correctly reported how agent-patient marks were used. Based on this information, an external researcher and I classified participants as *aware* or *unaware* of agent-patient marking with the help of a rubric (the transcriptions and the rubric can be found in Appendix C-7). Disagreements were discussed until a unanimous decision was made. Aware learners were those who could *at least* state that the noun ending in *–ak* was the agent/subject of the sentence or that the noun ending in *–a* was the object/patient of the sentence, for this already indicated that they knew that noun-marking was used to differentiate agents from patients. Unaware learners were those who were not able to identify the agent-patient marks or who were able to identify them but could not (correctly) say what conceptual role they marked. Next, awareness was transformed into a binary variable (1 = aware learner, 0 = unaware learner). I calculated the percentage of aware and unaware cognate and non-cognate learners. Finally, I noted whether aware learners reported noticing the target agent-patient marking during the exposure phase, which would suggest that learning had taken place in that phase and, thus, that it had not occurred during the testing phase.

3.6.5. Reading span task

I collected each participant's *partial reading span score*, defined as the total number of letters recalled in the correct order in each of the testing sets (Unsworth et al., 2005). For instance, if a participant correctly recalled 3 letters in a set of 4 letters and 3 letters in a set of 5 letters, the partial reading span score was 6. This score was automatically reported by the test. I chose the partial reading span measure instead of the *absolute reading span score* (the sum of all letters in perfectly recalled sets) because the former has more variance than the latter and allows for better differentiation between participants with various degrees of working memory capacity. A couple of Shapiro-Wilk tests indicated that the samples of reading span scores for cognate and non-cognate learners followed a normal distribution ($W = 0.98$, $p = .74$ and $W = 0.94$, $p = .11$, respectively). A Levene's test calculated with the function *leveneTest* from the *car* package (Fox & Weisberg, 2019) revealed that the variances of the two samples were equal ($F\ (1, 58) = 0.06$, $p = .81$). Because these assumptions were met, cognate and non-cognate learners' mean reading span scores were compared with a parametric two-sample t-test fitted with the function *t.test* from the *stats* package.

3.7. Results

3.7.1. First vocabulary-learning phase

3.7.1.1. Picture-word matching task

The picture-word matching task had to be performed until 100% accuracy. Cognate and non-cognate learners had to match pictures with (i) one of five non-cognate nouns or (ii) one of four either cognate or non-cognate verbs, respectively.

Picture-noun matching

The attempts necessary to match all nouns with their pictures ranged from 1-4 for cognate and non-cognate learners, with the exception of a non-cognate learner who had to perform the task a fifth time. Table 4.3 reports both groups' mean and median number of attempts at the task, dispersion measures and 95% confidence intervals for the two values. Both the mean and the median were slightly greater for non-cognate learners than for cognate learners. Yet, a Wilcoxon rank-sum test indicated that the distribution of number of attempts of non-cognate learners did not have significantly larger values than the distribution of number of attempts of cognate learners ($W = 349.5$, $p = .09$, $r = 0.22$).

	Cognate learners	Non-cognate learners
Mean	1.43	1.77
SD	0.77	0.97
95%CI	[1.14, 1.72]	[1.40, 2.13]
Median	1	2
MAD	0	1
95%CI	[1,1]	[1,2]

TABLE 4.3. Information regarding the number of attempts at the picture-noun matching task in Experiment 4 for cognate and non-cognate learners. This includes the mean, Standard Deviation (SD) and the 95% Confidence Interval (95%CI) of the mean, the median, the Median Absolute Deviation (MAD) and the 95%CI of the median.

Picture-verb matching

The attempts necessary to match all verbs with their pictures ranged from 1-3 for cognate learners and from 1-5 for non-cognate learners. Specifically, 60% (18/30) of cognate learners correctly matched all (cognate) verbs with their pictures in the first attempt at the task, 37% (11/30) did so in their second attempt and just one learner had to perform the task a third time. By contrast, only 43% (13/30) of non-cognate learners correctly matched all (non-cognate) verbs with their pictures in their first attempt, 33% (10/30) did so in a second attempt, 19% (3/30) needed a third attempt and the remaining four learners needed either a fourth attempt (2/30) or a fifth attempt (2/30). Table 4.4 summarises cognate and non-cognate learners' mean and median number of attempts at the task, dispersion measures and 95% confidence intervals for the two values. The mean and the median were greater for non-cognate learners than for cognate learners. A Wilcoxon rank-sum test revealed that the sample of number of attempts of non-cognate learners had marginally larger values than the sample of number of attempts of cognate learners, i.e. there was a cognate facilitation effect (W = 339.5, p = .07, small effect size of r = 0.23).

	Cognate learners	Non-cognate learners
Mean	1.43	2.00
SD	0.57	1.20
95%CI	[1.22, 1.65]	[1.55, 2.45]
Median	1	2
MAD	0	1
95%CI	[1,2]	[1,2]

TABLE 4.4. Information regarding the number of attempts at the first picture-verb matching task in Experiment 4 for cognate and non-cognate learners. This includes the mean, Standard Deviation (SD) and the 95% Confidence Interval (95%CI) of the mean, the median, the Median Absolute Deviation (MAD) and the 95%CI of the median.

3.7.1.2. Picture-naming task

The picture-naming task had to be performed until learners named all pictures of nouns and verbs correctly. Cognate and non-cognate learners named pictures of (i) the five non-cognate nouns and (ii) the four either cognate or non-cognate verbs, respectively.

Noun picture naming

Cognate and non-cognate learners named all pictures of nouns correctly in 1-3 attempts at the task, except for a cognate learner who performed the task four times. Table 4.5 shows both groups' mean and median number of attempts, dispersion measures and 95% confidence intervals for the two values. The mean number of attempts was larger for non-cognate learners than for cognate learners. However, the median of the two groups of learners was equal. A Wilcoxon rank-sum test indicated that the distributions of number of attempts of cognate and non-cognate learners had equal medians *and comparable shapes*. In other words, none of the distributions had significantly larger values than the other one ($W = 410.5$, $p = .54$, $r = 0.08$).

	Cognate learners	**Non-cognate learners**
Mean	1.97	2.07
SD	0.81	0.74
95%CI	[1.66, 2.27]	[1.79, 2.34]
Median	2	2
MAD	1	1
95%CI	[2,2]	[2,2]

TABLE 4.5. Information regarding the number of attempts at the noun picture-naming task in Experiment 4 for cognate and non-cognate learners. This includes the mean, Standard Deviation (SD) and the 95% Confidence Interval (95%CI) of the mean, the median, the Median Absolute Deviation (MAD) and the 95%CI of the median.

Verb picture naming

Cognate learners named all pictures of verbs correctly in 1-2 attempts at the task, but some non-cognate learners needed a third attempt. Specifically, while 83% (25/30) of cognate learners produced all verbs correctly in their first attempt, this was the case for just 30% (9/30) of non-cognate learners. The remaining cognate learners (17%, 5/30) named the verbs appropriately in their second attempt, but non-cognate learners had to attempt the task either twice (43%, 13/30) or three times (27%, 8/30). Table 4.6 shows cognate and non-cognate learners' mean and median number of attempts, dispersion measures and 95% confidence intervals for the two values. The mean and the median were larger for non-cognate learners than for cognate learners. A Wilcoxon rank-sum test indicated that the distribution of number of attempts of non-cognate learners had significantly larger values than the distribution of number of attempts of cognate learners, i.e. there was a cognate facilitation effect (*W* = 190, *p* < .001, large effect size of *r* = 0.56).

	Cognate learners	Non-cognate learners
Mean	1.17	1.97
SD	0.38	0.76
95%CI	[1.03, 1.31]	[1.68, 2.25]
Median	1	2
MAD	0	1
95%CI	[1,1]	[2,2]

TABLE 4.6. Information regarding the number of attempts at the first verb picture-naming task in Experiment 4 for cognate and non-cognate learners. This includes the mean, Standard Deviation (SD) and the 95% Confidence Interval (95%CI) of the mean, the median, the Median Absolute Deviation (MAD) and the 95%CI of the median.

3.7.2. Second vocabulary-learning phase

Cognate and non-cognate learners learnt the same four novel non-cognate verbs. Learners performed the picture-word matching task and the picture-naming task until they correctly matched all verbs with their pictures and they named all pictures of verbs appropriately.

3.7.2.1. Picture-word matching task

It took cognate and non-cognate learners 1-3 attempts to achieve 100% accuracy in this task. Both groups attempted the task the same mean and median number of times (Table 4.7). The samples of number of attempts of cognate and non-cognate learners had equal medians *and comparable shapes*. That is, none of the samples had significantly larger values than the other one (*W* = 431, *p* = .72, *r* = 0.05).

	Cognate learners	Non-cognate learners
Mean	1.37	1.37
SD	0.72	0.61
95%CI	[1.10, 1.63]	[1.14, 1.60]
Median	1	1
MAD	0	0
95%CI	[1,1]	[1,1]

TABLE 4.7. Information regarding the number of attempts at the second picture-verb matching task in Experiment 4 for cognate and non-cognate learners. This includes the mean, Standard Deviation (SD) and the 95% Confidence Interval (95%CI) of the mean, the median, the Median Absolute Deviation (MAD) and the 95%CI of the median.

3.7.2.2. Picture-naming task

Cognate and non-cognate learners attempted the task either once or twice. Table 4.8 shows the two groups' mean and median number of attempts, dispersion measures and 95% confidence intervals for the two values. While the mean number of attempts was larger for cognate learners than for non-cognate learners, the median of the two groups was equal. A Wilcoxon rank-sum test indicated that the distributions of number of attempts of cognate and non-cognate learners had comparable values (*W* = 510, *p* = .17, *r* = 0.18).

	Cognate learners	**Non-cognate learners**
Mean	1.23	1.10
SD	0.43	0.31
95%CI	[1.07, 1.39]	[0.99, 1.39]
Median	1	1
MAD	0	0
95%CI	[1,1]	[1,1]

TABLE 4.8. Information regarding the number of attempts at the second verb picture-naming task in Experiment 4 for cognate and non-cognate learners. This includes the mean, Standard Deviation (SD) and the 95% Confidence Interval (95%CI) of the mean, the median, the Median Absolute Deviation (MAD) and the 95%CI of the median.

3.7.3. Testing phase

3.7.3.1. Sentence-picture congruency task

<u>Performance on semantically incongruent sentence-picture pairs</u>

All semantically incongruent sentence-picture pairs had to be rejected (by contrast, semantically congruent sentence-picture pairs were syntactically congruent or incongruent and had to be either accepted or rejected, see next section). Cognate learners rejected a mean of 87.11% (*SD* = 33.54%, 95%CI = [84.36, 89.86]) of all semantically incongruent pairs with agent violation and a mean of 84.55% (*SD* = 36.18%, 95%CI = [81.59, 87.51]) of all semantically incongruent pairs with patient violation. Similarly, non-cognate learners rejected a mean of 85.49% (*SD* = 35.25%, 95%CI = [82.62, 88.37]) of all semantically incongruent pairs with agent violation and a mean of 81.50% (*SD* = 38.86%, 95%CI = [78.31, 84.69]) of all semantically incongruent pairs with patient violation. A generalized linear mixed effects model indicated that accuracy did not statistically differ as a function of Group of learners (β = 0.18, *SE* = 0.21, *z* = 0.89, *p* = .38), Type of sentence-picture pair (β = 0.23, *SE* = 0.16, *z* = 1.48, *p* = .14) or the interaction between the two (β = -0.11, *SE* = 0.26, *z* = 0.41, *p* = .68). The fact that both groups correctly identified a mismatch between the agent or the patient in the sentence and its associated picture in over 80% of the trials on average corroborates that they learnt the nouns very well. Likewise, the results corroborate that learning of nouns was comparable for cognate and non-cognate learners.

Performance on syntactically congruent and incongruent sentence-picture pairs

The structures as part of the linguistic system

Table 4.9 shows cognate and non-cognate learners' mean accuracy percentages when judging syntactically congruent SOV sentence-picture pairs (SOV_{congr}), incongruent SOV sentence-picture pairs ($SOV_{incongr}$), congruent OSV sentence-picture pairs (OSV_{congr}) and incongruent OSV sentence-picture pairs ($OSV_{incongr}$). These percentages are illustrated in Figure 4.7. As shown, cognate and non-cognate learners correctly judged significantly more than 50% of all congruent SOV and OSV sentence-picture pairs. However, they rejected less than 50% of all incongruent SOV and OSV sentence-picture pairs. Mean accuracy in incongruent OSV sentence-picture pairs was significantly below chance for both groups of learners (cognate learners, t (29) = -2.96, p < .01, d = -0.17; non-cognate learners, t (29) = -5.53, p < .001, d = -0.32). Mean accuracy in incongruent SOV sentence-picture pairs was also significantly below chance for cognate learners (t (29) = -2.16, p = .02, d = -0.13), but not for non-cognate learners (t (29) = -0.65, p = .26, d = -0.04).

		SOV_{congr}	**$SOV_{incongr}$**	**OSV_{congr}**	**$OSV_{incongr}$**
Cognate learners	*M*	88.11***	43.62*	85.57***	41.44**
	SD	32.42	49.68	35.20	49.35
	95%CI	[84.34, 91.89]	[37.79, 49.44]	[81.52, 89.63]	[35.75, 47.12]
Non-cognate learners	*M*	86.16***	48.07	73.21***	34.59***
	SD	34.59	50.05	44.36	47.65
	95%CI	[82.15, 90.16]	[42.23, 53.95]	[68.00, 78.43]	[29.10, 40.08]

TABLE 4.9. Cognate and non-cognate learners' mean accuracy (%), standard deviations (%) and 95% confidence intervals in syntactically congruent and incongruent SOV and OSV sentence-picture pairs in Experiment 4. *M* = Mean, *SD* = Standard Deviation, *95%CI* = 95% Confidence Interval. Significance from chance: * p < .05. ** p < .01. *** p < .001.

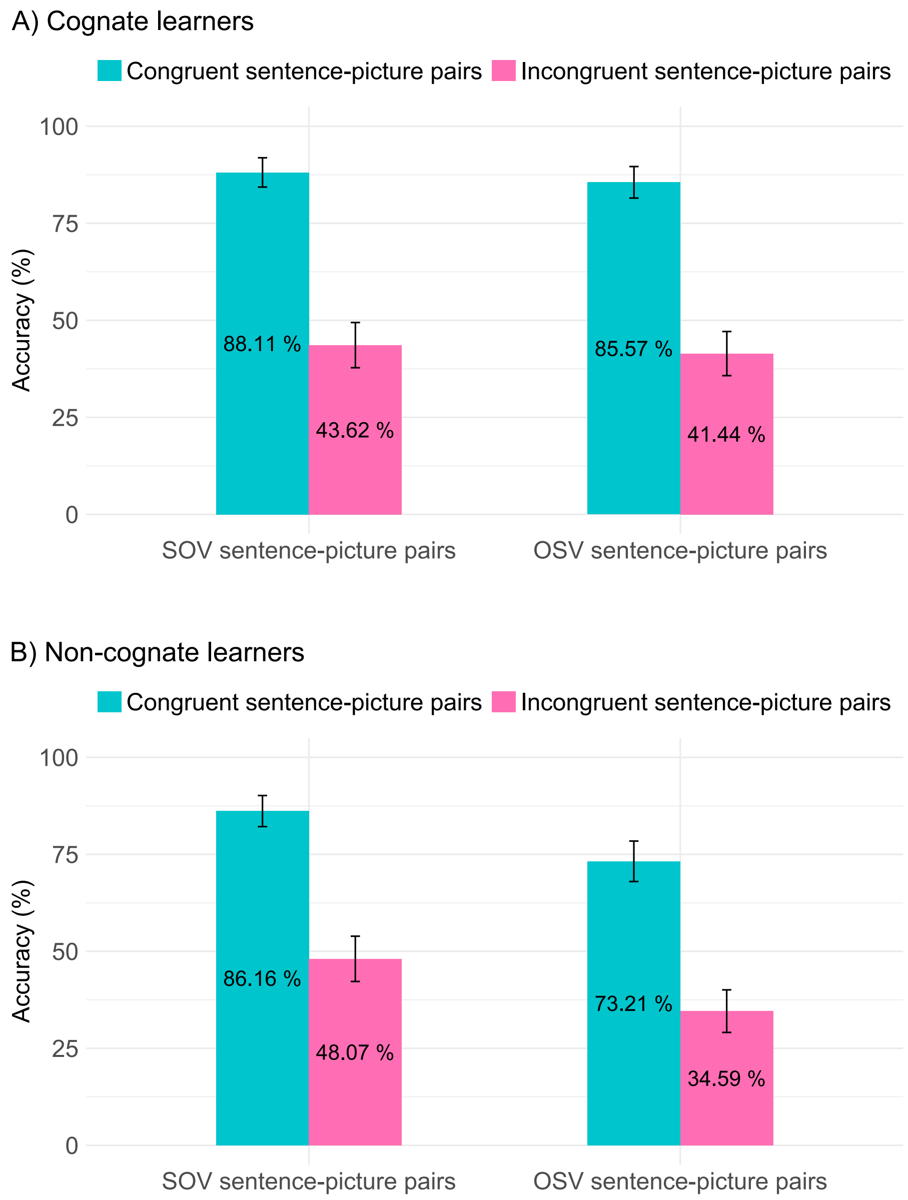


FIGURE 4.7. Mean accuracy (%) in syntactically congruent and incongruent SOV and OSV sentence-picture pairs for A) cognate learners and B) non-cognate learners in Experiment 4. Error bars represent 95% confidence intervals.

The descriptive analysis suggests that both groups of learners tended to judge most sentence-picture pairs as "correct". This was corroborated by the mean index of bias *c* measuring cognate and non-cognate learners' response bias when judging syntactically congruent and incongruent SOV and OSV sentence-picture pairs. In all cases, *c* was negative and significantly below zero (cognate learners, SOV_{congr} and $SOV_{incongr}$: M = -0.66, SD = 0.65; t (29) = -5.48, p <.001, d = -1; OSV_{congr} and $OSV_{incongr}$: M = -0.67, SD = 0.58; t (29) = -6.20, p <.001, d = -1.13. non-cognate learners, SOV_{congr} and $SOV_{incongr}$: M = -0.56, SD = 0.55; t (29) = -5.31, p <.001, d = -0.97; OSV_{congr} and $OSV_{incongr}$: M = -0.55, SD = 0.64; t (29) = -4.51, p <.001, d = -0.82). This tendency to accept SOV and OSV sentence-picture pairs irrespective of their syntactic congruency affected the two groups similarly, as indicated by a generalized linear mixed effects model testing for the interaction between Group of learners (Cognate vs. Non-cognate) and Word order of the sentence-picture pair (SOV vs. OSV) on accuracy. Specifically, the model yielded no effect of Group of learners (β = 0.35, *SE* = 0.42, *z* = 0.83; *p* = .41), no effect of Word order of the sentence-picture pair (β = 0.58, *SE* = 0.35, *z* = 1.63; *p* = .10) and no interaction between the two variables (β = -0.66, *SE* = 0.62, *z* = -1.07; *p* = .29).

Overall, these results provide no evidence that the SOV and the OSV structures with postpositional agent-patient marking were part of learners' linguistic system, for neither cognate nor non-cognate learners could distinguish between SOV and OSV sentence-picture pairs with correct and incorrect agent-patient marking. An explanation for this performance may be found in the strategies that participants reported following to conduct this task. This information was obtained in the debriefing phase, not in the testing phase. However, it is detailed below to facilitate the link between participants' strategies and their results in the sentence-picture congruency task.

Strategies used to perform the sentence-picture congruency task

Some participants reported conducting the sentence-picture congruency task using the agent-patient marking knowledge obtained during the exposure phase (40%, 12/30 of all cognate learners and 37%, 11/30 of all non-cognate learners). Others reported using their intuition (20%, 6/30 of all cognate learners and 27%, 8/30 of all non-cognate learners). Crucially, 40% (12/30) of all cognate learners and 37% (11/30) of all non-cognate learners reported using their vocabulary knowledge only, i.e. judging the congruency of sentence-picture pairs by only checking if the nouns in the sentence matched the ones in the picture or not. If the two nouns in the sentence coincided with the ones in the picture —as in syntactically congruent and incongruent pairs— that pair was judged as "correct". If one of the two nouns in the sentence did not match the ones in the picture —as in semantically incongruent pairs— that pair was judged as "incorrect". However, 42% (5/12) of these cognate learners and 18% (2/11) of these non-cognate learners verbalized the correct agent-patient marking in the debriefing phase and reported becoming aware of this marking during the exposure phase (see Section 3.7.4 below). This suggests that these participants learnt the target structures, but did not use their syntax knowledge to perform the congruency task. On the other hand, it could be that those learners who reported following this strategy but did not verbalize the agent-patient marking had unconscious knowledge of it. In sum, this means that for those participants following this strategy, syntax learning could not be captured in the congruency task. This diminishes the reliability of the results of this task, making it hard to conclude whether cognate and non-cognate learners learnt the target structures and whether one group learnt them significantly better than the other one.

3.7.3.2. Production task

The structures as part of the linguistic system

Table 4.10 shows a descriptive analysis of cognate and non-cognate learners' picture descriptions. First, it shows the percentage of SOV and OSV sentences written by the two groups of learners. To the right of this information is the percentage of SOV and OSV sentences with and without agent-patient marking. The rightmost column shows the percentage of SOV and OSV sentences with correct agent-patient marking. As shown, cognate and non-cognate learners wrote both SOV and OSV picture descriptions. This shows that they were aware that the agent/subject and the patient/object in the picture could appear in either the first or the second position of the sentence. Both groups of learners wrote more subject-initial sentences than object-initial ones and, descriptively, the proportion of SOV and OSV sentences in the two groups was similar. In addition, both cognate and non-cognate learners used agent-patient marking in significantly more than 50% of their SOV and OSV picture descriptions. This shows that, overall, both groups were aware that some marks had to be attached to the nouns in the sentence. Although descriptively, cognate learners wrote a much higher percentage of OSV sentences with agent-patient marking than non-cognate learners, the difference is only of eleven sentences, if we consider the total amount of OSV sentences written by the two groups. Finally, agent-patient marking was used correctly in significantly more than 50% of cognate and non-cognate learners' SOV and OSV picture descriptions with agent-patient marks, which suggests that the target structures, with verb-final word order and postpositional agent-patient marking, were part of learners' linguistic system.

Cognate learners	*SOV*	66.11% (158/239)	*With agent-patient marking*	63.92%*** (101/158)	*Correct agent-patient marking*	85.15%*** (86/101)
			Without agent-patient marking	36.08% (57/158)		
	OSV	33.89% (81/239)	*With agent-patient marking*	80.25%*** (65/81)	*Correct agent-patient marking*	86.15%*** (56/65)
			Without agent-patient marking	19.75% (16/81)		
Non-cognate learners	*SOV*	60.20% (121/201)	*With agent-patient marking*	62.81%** (76/121)	*Correct agent-patient marking*	68.42%*** (52/76)
			Without agent-patient marking	37.19% (45/121)		
	OSV	39.80% (80/201)	*With agent-patient marking*	66.25%** (53/80)	*Correct agent-patient marking*	62.26%* (33/53)
			Without agent-patient marking	33.75% (27/80)		

TABLE 4.10. Cognate and non-cognate learners' percentages of (i) SOV and OSV picture descriptions, (ii) descriptions with and without agent-patient marking and (iii) descriptions with correct agent-patient marking in Experiment 4. Significance from chance: * $p < .05$. ** $p < .01$. *** $p < .001$.

Comparing the establishment of the structures in cognate and non-cognate learners' linguistic system

A generalized linear mixed effects model tested whether cognate and non-cognate learners were comparably accurate when writing SOV and OSV sentences with agent-patient marking or, in other words, whether there was an interaction between Group of learners (Cognate vs. Non-cognate) and Word order of the picture description (SOV vs. OSV) on accuracy. The test revealed a significant effect of Group of learners (β = 3.92, SE = 1.89, z = 2.10, p = .036), no effect of Word order of the picture description (β = -0.30, SE = 0.57, z = -0.52, p = .60) and no interaction between the two variables (β = 0.43, SE = 1.15, z = 0.38, p = .71). The absence of a significant interaction indicates that the accuracy with which learners produced SOV vs. OSV sentences with agent-patient marking did not significantly vary as a function of group of learners. In addition, the lack of a significant effect of Word order suggests that, overall, cognate and non-cognate learners were not significantly more accurate when writing one or the other type of sentences. All this is evidence in favour that the SOV and the OSV structures with postpositional agent-patient marking were comparably established in the linguistic system and that this was the case for cognate and non-cognate learners, as predicted by the first claim of my hypothesis. I turn now to the significant effect of Group of learners. Figure 4.8 illustrates cognate and non-cognate learners' mean percentage of picture descriptions with verb-final word order (collapsing SOV and OSV) and correct agent-patient marking out of all picture descriptions with agent-patient marks. For cognate learners, the mean percentage of correct picture descriptions was 85.54% (SD = 35.27%, 95%CI = [80.14, 90.95]). For non-cognate learners, the mean percentage was lower, 65.89% (SD = 47.59%, 95%CI = [57.60, 74.18]). The significant effect of Group of learners reveals that, overall, cognate learners were significantly more accurate when writing verb-final picture descriptions with agent-patient marking than non-cognate learners. This supports the claim that the target structures were more robustly established in cognate learners' linguistic system than in non-cognate learners' linguistic system, as predicted by the second claim of my hypothesis.

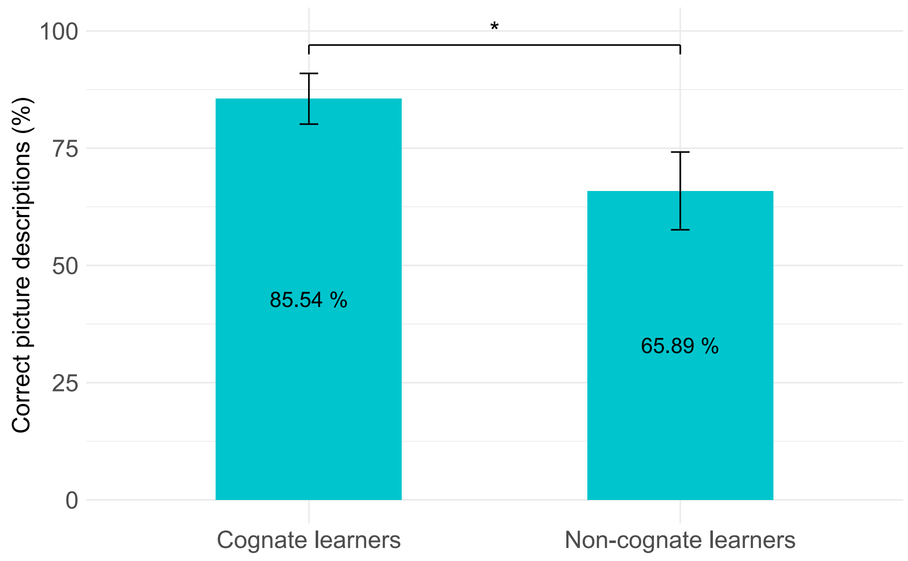


FIGURE 4.8. Mean percentage of picture descriptions with verb-final word order and correct agent-patient marking produced by cognate and non-cognate learners in Experiment 4. Error bars represent 95% confidence intervals.

3.7.4. Debriefing phase

In this section, I report participants' responses to the questions in the verbal report evaluating awareness of agent-patient marking. Participants' responses to the question about the strategies used to perform the sentence-picture congruency task were summarised when reporting the results of that task (see Section 3.7.3).

Most cognate learners (83%, 25/30) and non-cognate learners (87%, 26/30) reported noticing that nouns were different from their citation form when they appeared in a sentence. Out of these, 4% (1/25) of cognate learners and 15% (4/26) of non-cognate learners generally reported that some "suffixes", "letters" or "terminations" were attached to nouns. However, the majority of cognate learners (96%, 24/25) and non-cognate learners (85%, 22/26) reported that specific marks were added to the end of nouns. Starting with cognate learners, 50% (12/24) identified only one mark, which was *–ak* (P5, P10, P15, P23, P25, P26 and P27), *–k* (P11, P17, P21 and P28) or *–eak* (P14)[16]. The remaining 50% (12/24) of learners identified two or more marks, which were *–ak/–a* (P2, P3, P4, P8, P12, P16, P19, P29 and P30), *–k/–a* (P7), *–eak/–ea/–ia* (P9) or *–k/–ek* (P22). All but five of these 24 cognate learners correctly identified that a suffix ending in *k* marked the agent of the sentence and that a suffix not ending in *k* (if reported) marked the patient of the sentence. The remaining five learners did not report a function for these marks (P7, P12 and P21), reported that they were agreement marks (P16) or reported that the suffix ending in *k* marked the patient of the sentence (P22).

[16] P = participant.

Turning to non-cognate learners, within those learners who reported that specific marks were attached to nouns, 55% (12/22) identified just one mark, which was *–ak* (P15, P17, P19, P25, P26 and P27), *–k* (P21, P22, P24 and P30), *–a* (P29) or *–iak* (P11). The remaining 45% (10/22) of learners identified two or more marks, which were *–ak/–a* (P7, P14, P16 and P20), *–k/–a* (P13), *–eak/–ea* (P6), *–oak/–a/–e* (P28), *–eak/–ak* (P1), *–leak/–a* (P4) or *–urtzu/–aile* (P10). Twelve out of these 22 learners correctly reported that a mark ending in *k* signalled the agent of the sentence and that a mark not ending in *k* (if reported) identified the patient of the sentence. As for the remaining 10 learners, they did not report a function for these marks (P14, P17 and P22), reported that these were added "as a result of influence from the following word" (P29), reported that the suffix ending in *k* marked the patient of the sentence (P16) or said that these suffixes indicated the sentence's agent and patient, without specifying which marked each role (P1, P4, P10, P20 and P21).

In sum, 63% (19/30) of all cognate learners and 40% (12/30) of all non-cognate learners reported at least that a suffix ending in *k* was attached to the noun functioning as agent of the sentence and, thus, were considered *aware* of agent-patient marking. By contrast, the remaining 37% (11/30) of cognate learners and 60% (18/30) of non-cognate learners were considered *unaware* of this marking, for they either did not notice that nouns differed from their citation form when these appeared in a sentence or noticed that some marks were added to nouns but did not report how these marks were (correctly) used. All aware learners reported noticing the agent-patient marking during the exposure phase and they all wrote over 50% of correct picture descriptions in the production task, which evidenced syntax learning. By contrast, unaware learners did not produce more than 50% of correct picture descriptions.

3.7.5. Reading span task

Cognate learners' mean partial reading span score was 48.03 (*SD* = 10.35, 95%CI = [44.17, 51.90]). Similarly, non-cognate learners' mean score was 48.30 (*SD* = 10.08, 95%CI = [44.54, 52.06]). The difference between the two scores was not significant (t (58) = -0.10, p = .92, d = -0.03). This suggests that cognate and non-cognate learners had comparable working memory capacities and, hence, that the between-groups difference in syntax learning observed in the production task should not be attributed to variations in working memory.

3.8. Discussion

Experiment 4 showed for the first time that cognates might facilitate the acquisition of cross-linguistically dissimilar L2 structures by complete beginner L2 learners. More precisely, in this experiment I studied how Spanish natives without knowledge of Basque learnt two syntactic constructions based on Basque grammar —SOV and OSV structures with postpositional agent-patient marking— as a result of exposure to these structures in sentences containing two Spanish-Basque non-cognate nouns and either a Spanish-Basque cognate verb (cognate learners) or a non-cognate verb (non-cognate learners). First, I taught cognate learners and non-cognate learners the non-cognate nouns and either the cognate or the non-cognate verbs with the help of pictures. Then, I exposed learners to sentences exemplifying the L2 structures with the vocabulary learnt, each accompanied by a picture. Next, I taught learners novel non-cognate verbs and I tested learning of the structures using these verbs in a sentence-picture congruency task and in a written production task. The experiment ended with a verbal report.

Based on how I proposed that the MOGUL framework could explain how cognates are stored and processed in the linguistic system, together with how MOGUL explains acquisition by processing of cross-linguistically dissimilar L2 structures and the interaction between lexical and syntactic processing, I formulated my hypothesis. This hypothesis had two main claims. First, it claimed that since the two L2 structures were not present in learners' L1 and, thus, they both had to be acquired from input, they would end up being comparably established in learners' linguistic system. This would be the case when the structures were learnt with cognate verbs and with non-cognate verbs. Second, it claimed that when exposed to the target structures with cognate verbs as opposed to non-cognate verbs, the stronger activation of cognates compared to non-cognates would spread to the structures containing them, consequently causing these structures to become more robustly established in cognate learners' linguistic system than in non-cognate learners' linguistic system. The results of the production task supported these two claims. As evidence in favour of the first claim, cognate and non-cognate learners were comparably accurate when writing SOV vs. OSV picture descriptions with agent-patient marking and, overall, learners did not produce one of the two types of descriptions significantly more accurately than the other. As evidence in favour of the second claim, cognate learners wrote verb-final picture descriptions with agent-patient marking significantly more accurately than non-cognate learners did. In the following sections, I discuss the results of the (first) vocabulary-learning phase (focusing on the acquisition of cognate and non-cognate verbs), the testing phase and the debriefing phase.

3.8.1. Discussion of the first vocabulary-learning phase

In the first vocabulary-learning phase, learning was tested in a picture-word matching task and in a picture-naming task. Both tasks were performed until 100% accuracy, so that cognate and non-cognate learners mastered all (non-cognate) nouns and all (cognate or non-cognate) verbs to the same extent. The results of the picture-word matching task and the picture-naming task indicated that, while cognate and non-cognate learners learnt nouns in a comparable number of attempts, cognate verbs were learnt faster than non-cognate verbs (i.e. in significantly fewer attempts). The cognate facilitation effect observed in this experiment extends the findings of previous studies conducted with adults (de Groot & Keijzer, 2000; N. C. Ellis & Beaton, 1993; Lotto & de Groot, 1998; Marecka et al., 2021; Rogers, Webb, et al., 2015; Valente et al., 2018) and children (Antón & Duñabeitia, 2020; Comesaña et al., 2019; Comesaña, Soares, et al., 2012; Tonzar et al., 2009).

Like Experiment 4, these studies exposed learners to cognate and non-cognate words and then tested learning using one or more tasks. The exposure phase usually took one of two forms. In some studies, the L2 words were visually and/or aurally presented together with their L1 translations (Comesaña, Soares, et al., 2012; de Groot & Keijzer, 2000; N. C. Ellis & Beaton, 1993; Lotto & de Groot, 1998; Rogers, Webb, et al., 2015; Tonzar et al., 2009; Valente et al., 2018). In other studies, each L2 word was paired with a picture; words were presented in spoken and/or in visual form (Comesaña et al., 2019; Comesaña, Soares, et al., 2012; Lotto & de Groot, 1998; Marecka et al., 2021; Tonzar et al., 2009) and sometimes were embedded in a sentence (Antón & Duñabeitia, 2020). Experiment 4 combined these two forms of exposure, presenting L2 words with a picture and with their L1 translation. In most previous studies, words were displayed more than once during the exposure phase, e.g. twice (de Groot & Keijzer, 2000; N. C. Ellis & Beaton, 1993), three times (Comesaña et al., 2019; Lotto & de Groot, 1998; Tonzar et al., 2009), four times (Comesaña, Soares, et al., 2012) or more than four times (Marecka et al., 2021). In Experiment 4, each word was repeated four times. As for the tests measuring vocabulary learning, several studies used translation tasks (from L1 to L2, de Groot & Keijzer, 2000; N. C. Ellis & Beaton, 1993; Lotto & de Groot, 1998; Rogers, Webb, et al., 2015; Tonzar et al., 2009); from L2 to L1, Comesaña et al., 2019; Comesaña, Soares, et al., 2012; de Groot & Keijzer, 2000; N. C. Ellis & Beaton, 1993). Other studies used picture-naming tasks (Lotto & de Groot, 1998) and picture-word matching tasks (Antón & Duñabeitia, 2020; Marecka et al., 2021), like in Experiment 4. As mentioned, previous research has supported a cognate facilitation effect in L2 word learning, showing that cognates are translated (de Groot & Keijzer, 2000; N. C. Ellis & Beaton, 1993; Lotto & de Groot, 1998; Tonzar et al., 2009), named (Lotto & de Groot, 1998) and recognized (Marecka et al., 2021; Valente et al., 2018) faster and/or more accurately than non-cognates. In what follows, I discuss how I hypothesize that cognate and non-cognate verbs were processed during the vocabulary-learning phase and how this could have yielded the results observed, all within the MOGUL framework.

At the beginning of the vocabulary-learning phase, cognate and non-cognate learners saw pictures representing the nouns and the verbs one at a time. Each Basque noun or verb was written below its picture together with its Spanish translation and it was simultaneously played. Participants repeated each Basque noun or verb aloud and pressed the space bar to move on to the next word. Consider first how I hypothesize that non-cognate learners could have processed (and acquired by processing) a non-cognate verb such as *margotu* ("to paint"). I argue that the first time learners saw the picture representing the verb, a visual representation of this picture activated. The interface between the visual module and the conceptual module sought to coindex this visual representation with a conceptual representation (Sharwood Smith & Truscott, 2014, sec. 5.2.3., 5.4.3). At the same time, the learner read the L2 word *margotu* and its L1 translation, *pintar*. An orthographic representation of these words activated. The orthographic representation of the L1 word activated coindexed acoustic, phonological, syntactic and conceptual representations. The L1 conceptual representation was then coindexed with the visual representation of the picture. In addition, the L2 word was played; an acoustic and a phonological representation of this word activated and these were coindexed with the orthographic representation of *margotu*. To process the L2 word, its orthographic, acoustic and phonological representations had to be coindexed with a syntactic and a conceptual representation. I assume that, since the representations of its L1 translation were highly active, they won competition against any alternatives and were assigned the same index as the orthographic, acoustic and phonological representations of the L2 word. These newly established indices received a low resting activation level, which increased each time that the L2 word was processed (Sharwood Smith & Truscott, 2014, sec. 7.6.1).

Consider now how I hypothesize that cognate learners could have processed the cognate equivalent of the non-cognate verb *margotu,* i.e. *pintatu.* The process described for the non-cognate verb would be generally valid to explain processing of the cognate verb. However, some differences must be noted. First, when learners read the verb *pintatu,* not only the orthographic representation of this word activated, but also the orthographic representation of formally similar words, including its L1 translation, *pintar.* Similarly, when learners read the verb *pintar,* not only the orthographic representation of this word activated, but also the orthographic form of the L2 verb, due to their cross-linguistic similarity (Sharwood Smith & Truscott, 2014, sec. 3.3.5). Something similar occurred with the acoustic and phonological representations of *pintatu* and *pintar*: since the two words have similar pronunciations, the activation of the acoustic and phonological representations of one of the two words activated the acoustic and phonological representations of its translation. As a result, the orthographic, acoustic and phonological representations of *pintatu* and *pintar* were more strongly activated than those of *margotu* and *pintar*. The stronger activation of the representations of the L1 word spread to coindexed syntactic and conceptual representations, causing the chain of representations of the L1 verb to be more strongly activated when presented with an L2 cognate than with a non-cognate. This could have made it easier to coindex the L1's syntactic and conceptual representations with the L2's orthographic, acoustic and phonological representations. In familiar thinking terms, the cognate verb might have been processed and established in the linguistic system more easily than the non-cognate verb. In addition, each time that the verbs were subsequently processed during exposure, the co-activation of the L1 and the L2 cognate forms probably caused that the chain of representations of the L2 cognate verb was more strongly activated (i.e. had a higher current activation level) than the chain of representations of the equivalent non-cognate verb. Since a higher current activation level results in a higher resting activation level after processing, by the end of this exposure phase cognates should have had a higher resting activation level than non-cognates (Sharwood Smith & Truscott, 2014, sec. 3.3.5). Consequently, the former should have been more robustly established in the linguistic system than the latter. This would explain why cognate learners matched all verbs to pictures (picture-word matching task) and named all pictures of verbs (picture-naming task) until 100% accuracy in fewer attempts than non-cognate learners did.

Finally, the cognate facilitation effect observed in Experiment 4 could also receive an explanation outside of MOGUL. For instance, it could be explained by the Parasitic Model of vocabulary development (Ecke & Hall, 1998; Hall, 1996, 2002, and more). This model claims that the first step to learn an L2 (or L3) word is to establish a phonological and/or orthographic representation of it. On the one hand, if the learner's lexicon does not contain a similar L1 word form, as in the case of non-cognates, a new phonological and/or orthographic representation will be created. Then, this L2 form will be connected to the lemma of its L1 translation equivalent, linked, in turn, to a conceptual representation. The lemma and conceptual representations of the L1 word will be accessed when processing the L2 non-cognate. On the other hand, if the learner's lexicon contains an L1 word form similar to the one in the input, as in the case of cognates, the phonological and/or orthographic representation of the L1 word will be attached to a representation of the L2 word including only those features in which the two word forms differ, if any. When processing the L2 cognate, the lemma and conceptual representations of the L1 word will be accessed via the direct link between the L1 and the L2 word forms. Crucially, since most or all of the cognate form will already be established in the learner's mind, learning an L2 cognate will be easier than learning a non-cognate, for which a new formal representation has to be created. On the other hand, some researchers have tentatively interpreted the cognate facilitation effect within the Revised Hierarchical Model (RHM, Kroll & Stewart, 1994) or the Developmental Bilingual Interactive Activation model (BIA-d model, Grainger et al., 2010). These models assume that the initial stage of L2 vocabulary acquisition is the same (but differ in their account of later stages of acquisition). The two models propose that L1 and L2 forms are initially stored separately, but that conceptual representations for translation equivalents are shared. L1 word forms are directly linked to their conceptual representations. When an L2 word is first encountered, a direct connection is established between the L2 word form and its equivalent L1 word form. Meaning is accessed via this L1 form. The cognate facilitation effect in vocabulary acquisition could be explained by assuming that, due to cross-linguistic similarity, the connection between the L1 and the L2 cognate word forms is stronger than the connection between the L1 and the L2 non-cognate word forms. This would make access to meaning and, thus, processing and learning easier for the former than for the latter (Comesaña et al., 2019; Kroll et al., 1998, 2002; Sunderman & Schwartz, 2008; Valente et al., 2018).

3.8.2. Discussion of the testing phase

3.8.2.1. Sentence-picture congruency task

In the sentence-picture congruency task, cognate and non-cognate learners saw SOV and OSV sentence-picture pairs and had to judge whether a sentence and its paired picture were congruent or incongruent. Half of these sentence-picture pairs were semantically incongruent and included a mismatch between either the agent or the patient in the sentence and the picture. The other half were either syntactically congruent or incongruent. In the first case, the meaning expressed by the sentence matched the one expressed by the picture. In the second case, the sentence expressed the opposite agent-patient relationship of that expressed in the picture. Both groups of learners correctly rejected most semantically incongruent sentence-picture pairs. Performance was statistically similar across groups, confirming that as mentioned in the previous section, learning of nouns was comparable for cognate and non-cognate learners. By contrast, while both groups judged most syntactically congruent sentence-picture pairs as required, this was not the case for syntactically incongruent pairs. This could evidence that the two target structures, with verb-final word order and postpositional agent-patient marking, were not established in learners' linguistic system. Yet, this was not necessarily true.

Learners were biased towards judging syntactically congruent and incongruent sentence-picture pairs as congruent, and this was at least partly the result of some participants' conscious strategy to judge sentence-picture pairs according to the match or mismatch between the vocabulary in the sentence and the picture. Since in semantically incongruent sentence-picture pairs one of the nouns in the sentence never matched the ones in the picture, most pairs were judged as incongruent. However, since in syntactically congruent and incongruent sentence-picture pairs the nouns in the sentence always matched the ones in the picture, most pairs were judged as congruent. As mentioned in Section 3.7.3, this bias indicates that the results of the task are not truly representative of cognate and non-cognate learners' learning of the structures. This bias is also quite unexpected given the characteristics of the materials of the sentence-picture congruency task. That is, 75% of all sentence-picture pairs were incongruent and only 25% of all sentence-picture pairs were congruent. This could have caused participants to have an overall bias towards judging most sentence-picture pairs as incongruent. Yet, this was not the case; both groups of learners tended to judge as incongruent or "incorrect" most semantically incongruent sentence-picture pairs, but they were biased towards judging as congruent or "correct" most syntactically congruent and incongruent sentence-picture pairs. Next, I comment on how learners could have processed the structures during the exposure phase and how this could have yielded the results obtained. Since the response bias observed affected cognate and non-cognate learners equally, I will discuss this for the two groups generally.

I will start by reviewing how I hypothesized that Spanish natives without knowledge of Basque would learn the two structures (cf. Section 2.2). I argued that, when learners processed (SOV or OSV) sentences, the syntactic processor would create a syntactic representation matching the input, e.g. [$_{CP}$ NP [$_{VP}$ NP V]]. The two NPs, consisting of a noun and a suffix (*–ak* or *–a*), would be processed as whole forms, since a compositional form would be non-transparent, i.e. there would be no syntactic and conceptual representations associated with *–ak* and *–a* in the linguistic system and it would not be obvious what these representations were. Due to the agent-first preference, the first NP would be interpreted as the agent of the sentence and the second NP would be interpreted as the patient. This would lead to a correct interpretation of SOV sentences, but not of OSV sentences. That is, learners would interpret OSV sentences as if they were subject-initial, yet, the picture accompanying them would not support this interpretation. After a few times misprocessing OSV sentences, learners would realize that the language had flexible word order and that the first noun of the sentence was not necessarily the agent. This would cause that, as more SOV and OSV sentences were processed, the conceptual roles of AGENT and PATIENT were coindexed with the suffixes *–ak* and *–a*, respectively, leading to the creation of the target syntactic structures with verb-final word order and postpositional agent-patient marking. These structures would initially receive a low resting activation level, which would increase each time that they proved useful for processing.

I argue that overall cognate and non-cognate learners processed SOV and OSV sentences as hypothesized, since Experiment 4 provided evidence that the two groups learnt the target structures (cf. production task). However, in light of the results of the sentence-picture congruency task, I also argue that for some learners processing may have differed in various ways. First, it could be that when processing sentences during the exposure phase, some learners realized that word order was not a reliable cue to agentivity, but did not get to process the NPs in SOV and OSV sentences as compositional forms. This may be because they needed to misprocess a larger number of OSV sentences before the coindexing of AGENT and PATIENT with *–ak* and *–a* was triggered*.* In this case, the syntactic representation of the sentences in these learners' linguistic system would have included two whole-word representations of NPs followed by a verb, with the syntactic representation of the two NPs coindexed with the conceptual roles of both AGENT and PATIENT. This would capture that the agent and the patient could appear in sentence-initial position, but indistinctly. Thus, when exposed to sentence-picture pairs during the congruency task, learners would have had no way of identifying when the first noun of the sentence was the agent and when it was the patient and, hence, whether the sentence and its associated picture were congruent or not. This could have caused that these learners judged all SOV and OSV sentence-picture pairs as congruent, yielding the response bias observed.

Second, it could be that some learners processed SOV and OSV sentences as hypothesized, but that the coindexing of the suffixes *–ak* and *–a* with the appropriate conceptual roles occurred towards the end of the exposure phase. This could have caused that there were not enough processing opportunities for the target structures to achieve high resting activation levels. In this case, it is possible that these structures co-existed in the linguistic system with a non-target structure including two whole-word representations of the NPs coindexed with the conceptual roles of both AGENT and PATIENT. If the coindexing of suffixes with conceptual roles occurred towards the end of exposure, the latter (non-target) syntactic representation could have had a higher resting activation level than the former (target) syntactic representations. In this light, when exposed to sentence-picture pairs during the congruency task, the low resting activation level of the target structures could have caused that the non-target structure oftentimes was selected to process the sentences. In familiar thinking terms, learners could have learnt the structures, but learning could have been too weak for them to be used during the congruency task. Consequently, knowing that subject-initial and object-initial sentences were both possible in the language, learners could have opted to judge all sentence-picture pairs as congruent.

Finally, it could be that some learners processed sentences as hypothesized, but that they found the sentence-picture congruency task too challenging. As a result, the syntactic processor could have turned to a parsing strategy by which it ignored the agent-patient marks in the input. That the congruency task could have been too challenging is not unreasonable if we consider that learners were presented with a sentence-picture pair for approximately a second (the duration of the audio of the sentence) and then were given 5 seconds to decide whether the sentence was congruent with the picture or not. In this short amount of time, learners had to (i) process the sentence, which involved accessing lexical items, integrating them into a syntactic structure and deriving an interpretation of the sentence, (ii) access a conceptual representation of the picture and (iii) compare the two to decide whether the sentence and the picture were semantically and syntactically congruent. The lexical items and the structures were learnt just before the congruency task and the exposure leading to learning was relatively brief. Hence, processing sentences might have been overall quite costly and, consequently, performing the congruency task in the time given could have been too demanding. As explained in Chapter 2, the MOGUL framework, in line with other approaches to L2 acquisition, claims that processing a sentence requires not only global awareness of the input (in this case, activating an auditory or visual representation of a sentence) but also *noticing* the structure in the input (in this case, activating the target syntactic representation of a sentence). This syntactic representation then becomes *intake* and is processed by the syntactic processor (Sharwood Smith & Truscott, 2014, sec. 9.3.2). I propose that learners who found the congruency task too challenging may have been globally aware of the input, which included the agent-patient marks, but could have activated a syntactic representation of the SOV and OSV sentences that did not include these agent-patient marks, i.e. these did not become intake.

The reasoning described could also be compatible with the Lexical Bottleneck Hypothesis (Hopp, 2018). As mentioned in Section 1.2, this hypothesis claims that lexical processing occurs before syntactic processing and has a direct influence on it. Specifically, if lexical processing is very costly, it may use up all the resources required to carry out a target-like syntactic parse. Based on this, it could be that processing the lexical items in a sentence during the congruency task was so costly that there were not enough resources to achieve a target syntactic processing and, as a result, that the processor turned to the parsing strategy described. In any case, if the sentences in the sentence-picture congruency task were processed without attending to the agent-patient marks, then it could be that learners made their congruency judgements based on whether or not the lexical items in the sentences matched the ones in the picture, which would have led them to accept syntactically congruent and incongruent sentence-picture pairs. This is the strategy that some learners reported following in the verbal report.

3.8.2.2. Production task

In the production task, cognate and non-cognate learners saw pictures of transitive actions and had to write a sentence that described each picture using the nouns and verbs learnt. As mentioned, this task provided evidence that learning of the target structures was greater for cognate learners than for non-cognate learners. Specifically, the former wrote a significantly larger number of sentences with verb-final word order and correct agent-patient marking than the latter. In addition, overall learners wrote a comparable number of SOV and OSV sentences with correct agent-patient marking. While learners did not always use agent-patient marking in their descriptions, both cognate and non-cognate learners wrote subject-initial and object-initial sentences. Both groups wrote more subject-initial sentences than object-initial ones and the proportion of the two types of sentences was similar across groups. In the previous section, I proposed that by the end of the exposure phase, SOV and OSV sentences could have been represented in learners' linguistic system in different ways. In what follows, I discuss how the representation of these sentences could have varied for cognate learners vs. non-cognate learners and how this could have yielded the findings observed.

I first review how I hypothesized that learners would process (SOV or OSV) sentences with cognate and non-cognate verbs when first exposed to them (cf. Section 2.2). In brief, as mentioned, I hypothesized that upon reading and listening to the target sentences, a verb-final syntactic representation would be created and the two NPs in these sentences would be processed as whole-word forms. Following the agent-first preference, SOV and OSV sentences would both be interpreted as agent/subject-initial. The mismatch between learners' interpretation of OSV sentences and their accompanying picture would eventually lead to coindexing the conceptual roles of AGENT and PATIENT with the forms *–ak* and *–a*. The target syntactic structures would be created and, with each processing opportunity, their resting activation level would increase. Regarding the role of cognates, I hypothesized that cognate verbs would be processed faster than non-cognate verbs, since the former would have a higher current and resting activation level than the latter and, thus, would be activated and selected for processing faster. This would cause that the sentences containing a cognate verb were processed faster than the sentences containing a non-cognate verb and, consequently, that the match or the mismatch between the interpretation of the sentence and its accompanying picture was detected faster for the first type of sentences than for the second. I argued that detecting the misinterpretation of OSV sentences as SOV led to learning agent-patient marking; hence, I hypothesized that cognates would facilitate the process by which this learning was triggered.

On the other hand, I hypothesized that cognates would lead to greater learning of the target structures than non-cognates. This hypothesis was based on the assumption that, if the chain of representations constituting a cognate had a higher current activation level than the chain of representations constituting a non-cognate, then stronger activation would spread from the syntactic representation of the cognate verb than from the syntactic representation of the non-cognate verb to the structure containing it. As a result, the target structures would have a higher current activation level when processed with a cognate than with a non-cognate. This would cause that the resting activation level of the structures processed with cognates became higher than that of the structures processed with non-cognates after each processing and, thus, that by the end of the exposure phase, the former were more robustly established in learners' linguistic system than the latter. The results of the production task confirmed this hypothesis. That is, since learning of the structures was greater for cognate learners than for non-cognate learners, the former were significantly more accurate than the latter when writing picture descriptions with verb-final word order and agent-patient marking. These results and the reasoning behind them go in line with some of the findings of Chapter 3. In that chapter, I argued that the stronger activation of high frequency verbs compared to low frequency verbs facilitated the acquisition of an L1-L2 dissimilar structure. Likewise, in Experiment 4 I argue that the stronger activation of cognate verbs compared to non-cognate verbs facilitated the acquisition of L1-L2 dissimilar structures.

Notably, the facilitative role of cognates in L2 syntax acquisition could also be tentatively explained by the declarative/procedural model of second language acquisition (Morgan-Short & Ullman, 2011; Ullman, 2001a, 2001b, 2004, 2008, 2012) and by the Lexical Bottleneck Hypothesis (Hopp, 2018), if it were applied to L2 acquisition. According to the declarative/procedural model, content words are learnt by the declarative memory system after a brief exposure and grammatical structures are usually learnt by the procedural memory system after a longer exposure. Once grammatical structures are robustly learnt by the procedural system, they are automatized and can be used productively. I proposed that processing sentences with cognates as opposed to non-cognates during the exposure phase should have led to a faster detection of the match or mismatch between a sentence and its paired picture. This could have caused that the coindexing of *–ak* and *–a* with the conceptual roles of AGENT and PATIENT and the creation of the target structures occurred faster (i.e. in fewer trials) for cognate learners than for non-cognate learners. In terms of the declarative/procedural model, cognate learners' procedural memory system would have had more opportunities (more trials) to process and, thus, consolidate or automatize the newly created structures than non-cognate learners' system. Turning to the Lexical Bottleneck Hypothesis, it argues that a costly (e.g. slower) lexical processing may deplete the resources needed to perform a target-like syntactic computation. Therefore, the hypothesis predicts, just as I proposed, that the slower processing of non-cognates compared to cognates could have led to slower sentence processing and, hence, to slower detection of the match or mismatch between the interpretation of a sentence and its picture. In addition, it predicts that processing sentences with non-cognates, as opposed to cognates, may have been so costly that it depleted the resources necessary to subsequently learn agent-patient marking, causing learning of the structures to be greater for cognate learners than for non-cognate learners.

In the previous section, I also mentioned the possibility that by the end of the exposure phase not all learners had learnt the target structures, or not as hypothesized. Specifically, I discussed that some learners could have realized that the mini-language had flexible word order and that the first noun in a sentence was not necessarily the subject or agent, but that input could have been insufficient for them to coindex *–ak* and *–a* with the conceptual roles of AGENT and PATIENT. The fact that these learners did not learn how agent-patient marking was realized would explain why cognate and non-cognate learners did not always use agent-patient marking in their picture descriptions. Additionally, I discussed the possibility that the target structures (with the appropriate word order and agent-patient marking) were established in learners' linguistic system but that these structures were created towards the end of exposure and, thus, that they had a very low resting activation level. In this case, learners would have learnt the structures but learning would have been so weak that they could have chosen not to use agent-patient marking in the production task.

Finally, regardless of whether cognate and non-cognate learners learnt the structures as I had hypothesized or not, I assume that all learners learnt that the mini-language allowed both the subject or agent and the object or patient to appear in sentence-initial position. This would explain why cognate and non-cognate learners wrote SOV and OSV picture descriptions irrespective of whether agent-patient marking was used or not, i.e. sentences in which the subject/agent or the object/patient in the picture was the first word. Both cognate and non-cognate learners preferred to write subject-initial sentences over object-initial sentences, and in a similar proportion. This could be explained by the agent-first preference, i.e. the preference in most languages of the world (including Spanish, learners' L1) to put agents before patients (Dryer, 2013). Additionally, I hypothesized that neither the SOV nor the OSV structure with postpositional agent-patient marking would be more robustly established in cognate and non-cognate learners' linguistic system than the other one. This was because the two structures were dissimilar to L1 grammar and, thus, both had to be learnt from scratch. This would explain why neither of the two groups of learners wrote significantly more SOV than OSV picture descriptions with correct agent-patient marks.

3.8.3. Discussion of the debriefing phase

Experiment 4 used an implicit learning paradigm, i.e. learners were not informed that they would be exposed to two structures during the exposure phase nor that they had to learn two constructions. Implicit learning paradigms usually result in implicit or unconscious (non-verbalizable) knowledge of the learning target (e.g. Kim & Fenn, 2020; Leung & Williams, 2006; Rebuschat, 2009; Tagarelli et al., 2016; Williams, 2005). Nevertheless, as mentioned in Section 3.4.5, despite learners not being instructed to look for patterns in the input during exposure, they could have consciously or unconsciously done so, since they were instructed to look at each picture, and listen to and read the sentence accompanying it. This is especially likely if we take into consideration that L2 learners have metalinguistic awareness. Because of this, I expected that the implicit learning paradigm resulted in explicit or conscious (verbalizable) knowledge of agent-patient marking, at least for some learners. This is not unprecedented, for, as mentioned in Chapter 1 (Section 1.3), incidental exposure may result in explicit knowledge for some learners and implicit knowledge for others (e.g. Kim & Fenn, 2020; Robinson, 1997; Williams, 2005).

Around half of cognate and non-cognate learners who reported that specific marks were added to the end of nouns verbalized two marks. Most learners reported marks that ended in *k* and *a*. The other half of cognate and non-cognate learners verbalized just one mark, which could have various forms, but in all but one case ended in *k*. This suggests that the mark–*ak* was much more salient than the mark –*a*, both for cognate and non-cognate learners. A possible explanation for this could lie in the characteristics of word-final codas in Spanish, learners' native language. Specifically, while nouns can and very often end in the letter *a* (or the sound [a]) in Spanish, finding the letter *k* (or the sound [k]) in word-final position is rare and only occurs in borrowings, e.g. *anorak, kayak* (Colina, 2009). Considering this, nouns ending in *k* could have attracted learners' attention throughout the experiment, increasing the likelihood of recalling a suffix ending in *k* in the debriefing phase.

More than 60% of cognate learners correctly reported, at least, that a mark ending in *k* was added to the noun acting as agent or subject of the sentence, but less than half of non-cognate learners did so. Hence, the experiment resulted in conscious knowledge of agent (-patient) marking for the majority of cognate learners, but not for the majority of non-cognate learners. All aware learners produced more than 50% of correct picture descriptions in the production task, which shows syntax learning. These learners indicated that they had become aware of agent-patient marking during the exposure phase, which suggests that the structures were learnt during that phase and not during the testing phase. Unaware learners did not verbalize how the agent-patient marks were (correctly) used and did not write more than 50% of correct picture descriptions, which could indicate that these learners did not learn the target structures. Another possibility is that unaware learners learnt the structures despite this not being reflected in the production task, but that the knowledge derived from this learning was unconscious and, thus, not verbalizable. Alternatively, it could be that learners did not have the confidence or the ability to express their knowledge with words. A way to differentiate between these options could have been to look at unaware learners' performance in the sentence-picture congruency task. If their accuracy when judging syntactically congruent and incongruent SOV and OSV sentence-picture pairs had been above chance, this would have indicated that the structures were part of learners' linguistic system. However, since the results of the congruency task are unreliable, it is not possible to determine whether unaware learners did not learn the structures or learnt them, but unconsciously.

4. Cognates and the acquisition of cross-linguistically similar L2 structures

Experiment 4 showed that cognates facilitated learning of two cross-linguistically dissimilar L2 structures exemplifying the verb-final word order and postpositional agent-patient marking in the L2. In the second part of Chapter 4, I report Experiment 5, which investigated whether and how the facilitative role of cognates varied when the L2 structures to be learnt embodied a word order and agent-patient marking similar to the L1. As in the first part of the chapter, I first introduce the structures studied in Experiment 5 and I consider how cognates could influence the acquisition of such structures before presenting the experiment and discussing its results.

4.1. Cross-linguistically similar structures studied in Experiment 5

The object of study of Experiment 5 were two structures made up of a Basque vocabulary but having verb-medial word order and prepositional patient marking, like Spanish. The patient marking was the same in the two structures, but the word order alternated between SVO (9a) and OVS (9b). In these structures, the subject was always an agent (A) and the object was always a patient (P).

(9) a. Antzezle pintatu a gidari.
actor.A paint P pilot
"The actor is painting the pilot."

b. A gidari pintatu antzezle.
P pilot paint actor.A
"The actor is painting the pilot."

(Example sentences used in Experiment 5)

The sentences in (9) express the same meaning. In (9a), the sentence starts with an animate noun (the agent of the action denoted by the verb) and bears no overt agent mark. This noun is followed by a verb, the word *a* and another animate noun, with *a* being a patient mark, i.e. indicating that the following noun is the patient of the verb. In (9b), the order of the agent and the patient is reversed. Since in principle both *antzezle* ("actor") and *gidari* ("pilot") could be the agent or the patient of *pintatu* ("paint"), to interpret the reversible sentences in (9) correctly, it is necessary to pay attention to the patient mark. This word order and agent-patient marking resemble the ones in Spanish, although some differences must be noted. First, the Basque verbs in the sentences in (9) are in citation form, although they express the meaning of a conjugated verb[17]. By contrast, in Spanish SVO (10a) and OVS (10b) sentences contain a conjugated verb that is formally different from its citation form. Second, in Spanish singular countable nouns are preceded by a determiner, but the nouns in Experiment 5 were not. I decided to use verbs in citation form and no determiners because, as in Experiment 4, in Experiment 5 participants learnt bare nouns and verbs (see Procedure, Section 5.4). Including articles or a conjugated verb in the experimental sentences would have meant introducing new elements to be learnt other than the word order and patient marking, complicating the materials unnecessarily. Finally, in Spanish the preposition (PREP) *a* precedes the animate and specific direct object of a transitive verb. When the object is left-dislocated, as in OVS sentences, the preposed argument is typically in a relation of co-reference with a clitic pronoun preceding the verb, e.g. in (10b) the object/patient *paciente* ("patient") correlates with the accusative (ACC), third person singular (3SG) clitic pronoun (CL) *lo* ("him"). Yet, this clitic is not necessarily used in all varieties of Spanish (Fábregas, 2013). In Experiment

[17] Remember that this is not the case in natural Basque, where the citation form of the verb differs from its conjugated form (cf. Section 2.1).

5, I decided not to include the clitic pronoun in OVS sentences because I wanted the patient to be denoted by a single mark irrespective of word order.

(10) a. La enfermera salvó al paciente.
The nurse.NOM.A save $_{\text{PST.3SG}}$ PREP.D.P.ACC patient
"The nurse saved the patient."

b. Al paciente lo salvó la enfermera.
PREP.D.P.ACC patient CL.3SG.ACC save $_{\text{PST.3SG}}$ the nurse.NOM.A
"The nurse saved the patient."

(Gutiérrez, 2008, p. 370)

Finally, although the sentences in (9) are made up of Basque words, they differ from Basque in several aspects. First, while Basque admits nearly all constituent combinations, including SVO and OVS, its canonical word order is SOV (Aldezabal et al., 2003; De Rijk, 1969; Pastor, 2019, 2020) and the verb-medial structures are much less frequent than their verb-final counterparts. For example, of all transitive sentences in the *Ereduzko Prosa Gaur* corpus (Sarasola et al., 2009)[18], 52% are SOV and just 26% are SVO. Likewise, 17% of the sentences are OSV, but just 2% are OVS (Pastor, 2020). In addition, Basque has postpositions instead of prepositions and, as shown in Section 2.1, the agent of transitive sentences is marked with a –*k,* while the patient of transitive sentences is morphologically unmarked. This contrasts with the sentences in (9), in which the patient of the transitive verb is marked with the preposition *a* and the agent of the transitive verb is not marked.

4.2. The influence of cognates on the acquisition of cross-linguistically similar L2 structures

In this section, I discuss how I propose that having a Spanish-Basque cognate verb as opposed to a non-cognate verb in the cross-linguistically similar structures described in the previous section could affect their acquisition by Spanish natives with no knowledge of Basque. As usual, I will mostly concentrate on acquisition by processing in MOGUL's syntax module, commenting on the interaction with the acoustic, visual, phonological and conceptual modules when needed. First, I will focus on how Spanish natives would process the SVO sentences with prepositional patient marking in (11) from the moment they are first exposed to them, assuming that they have previously learnt the nouns and verbs in the sentences, as in Experiment 5 (see Section 5.4). In (11a), the verb is cognate between Spanish and Basque and in (11b), it is non-cognate in the two languages.

(11) a. Antzezle pintatu a gidari.
actor.A paint P pilot
"The actor is painting the pilot."

[18] The corpus contains 25.1 million words; 13.1 million are taken from 287 books and 12 million are taken from newspaper articles published in Spain (Berria) and France (Herria).

b. Antzezle margotu a gidari.
actor.A paint P pilot
"The actor is painting the pilot."

I assume that when learners read or listen to *antzezle,* an orthographic and/or phonological representation of this word will activate. These representations will activate a coindexed syntactic representation (e.g. [N_m]) which, in turn, will activate larger syntactic representations in which it occurs. Since learners will have never encountered this word in a sentence, the representations activated will be those of the L1, including [$_{CP}$ NP [$_{VP}$ V]], [$_{CP}$ NP [$_{VP}$ V PP]] and [$_{CP}$ NP [$_{VP}$ V NP]]. These will compete for selection as input is subsequently processed. If one of them proves adequate, no new, L2-specific representation will be constructed (Sharwood Smith & Truscott, 2014, sec. 4.2). In addition, since in Spanish singular countable nouns are preceded by a determiner, a representation such as [$_{NP}$ D N_m] may also activate. As described for Experiment 4 (Section 2.2), I propose that the syntactic representation of the determiner may be coindexed with a null orthographic and/or phonological representation that increases in resting activation level each time that an NP without an overt determiner is encountered. I assume this for the two NPs in the sentence.

Meanwhile, the (cognate or non-cognate) verb will be processed. Activation will spread from its orthographic/phonological representation to the coindexed syntactic representation (e.g. [V_n]) and to larger syntactic representations containing it, further raising the current activation level of [$_{CP}$ NP [$_{VP}$ V]], [$_{CP}$ NP [$_{VP}$ V PP]] and [$_{CP}$ NP [$_{VP}$ V NP]]. These constructions, acquired as part of the L1, will contain a finite verb, but the verb that learners encounter will be in its citation form. I propose that the orthographic/phonological and conceptual representations of the L2 verb may be coindexed with the syntactic representation of the finite verb in the L1 constructions, allowing learners to interpret that in this language, non-finite and finite verbs have the same surface form, as it occurs in other languages (e.g. English, *to eat – I eat*). Finally, *a gidari* will be processed. The orthographic/phonological form of *a* will activate coindexed syntactic representations. In Spanish, *a* is a preposition, usually followed by an NP. Thus, a syntactic representation such as [$_{PP}$ $PREP_q$ NP] will activate. Additionally, *gidari* will activate the syntactic representation of a noun and its activation will further spread to the [$_{PP}$ $PREP_q$ NP] representation. This will cause that *a gidari* is processed as a PP, which, in turn, will spread its activation to the larger syntactic representation active, [$_{CP}$ NP [$_{VP}$ V PP]]. To interpret the sentence, conceptual roles will be assigned to the verb's arguments. As mentioned in Section 2.2, there is a general tendency to interpret the first animate NP in a sentence as the agent (*agent-first preference*) and this is also true for Spanish natives. In addition, in Spanish *a* precedes the animate and specific object/patient of a transitive verb. Considering all this, I assume that the first NP (*antzezle*) will be correctly interpreted as the agent of the sentence and the second NP (*gidari*), as the patient. Since the sentence will be successfully processed, the resting activation level of its syntactic representation will increase and, each time that a sentence with this structure is processed, its resting activation level will increase even more (Sharwood Smith & Truscott, 2014, sec. 3.3.5, 4.2).

As for the influence that the cognate or non-cognate verb will have on processing, this could be small or even negligible. On the one hand, the formal similarity between the cognate and its Spanish translation should cause that the cognate has a higher current activation level than the non-cognate. Consequently, the activation spreading from the syntactic representation of the verb to the syntactic representation of the structure should be higher for the former than for the latter. This should cause that the structure processed with the cognate (11a) has a higher current activation level than the one processed with the non-cognate (11b) and, thus, a higher resting activation level when processing terminates. On the other hand, the [$_{CP}$ NP [$_{VP}$ V PP]] structure is assumed to be stored in the linguistic system by the time L2 acquisition starts and to have a high resting activation level due to its extensive use in the L1. The increases in resting activation level derived from processing are large when experience with the input is low but, as the input's resting activation level becomes higher, the increases diminish until, at some point, the resting level increases no more (Sharwood Smith & Truscott, 2014, sec. 4.6.5). Considering this, additional processing of the [$_{CP}$ NP [$_{VP}$ V PP]] structure should lead to small or no increases in its resting activation level. Consequently, the difference in resting level derived from processing the structure with a cognate vs. a non-cognate verb may be small or even imperceptible.

I now turn to how I hypothesize that Spanish natives with no knowledge of Basque would process the equivalent OVS sentences with prepositional patient marking in (12) from first exposure, assuming that they are familiar with the nouns and verbs in these sentences. (12a) contains a Spanish-Basque cognate verb and (12b) contains the corresponding non-cognate verb.

(12) a. A gidari pintatu antzezle.
P pilot paint actor.A
"The actor is painting the pilot."

b. A gidari margotu antzezle.
P pilot paint actor.A
"The actor is painting the pilot."

I propose that when learners read or listen to *a gidari* for the first time*,* two syntactic representations may activate and compete for selection by the syntactic processor. The representations activated will be those acquired as part of the L1. On the one hand, *a gidari* could activate a syntactic representation such as [$_{PP}$ PREP NP] for, as mentioned, in Spanish *a* is a preposition and it is often followed by an NP. In Spanish singular countable nouns are preceded by a determiner, which, in the case of *a gidari,* could be orthographically/ phonologically null. On the other hand, the orthographic and/or phonological representations of *gidari* could activate the coindexed syntactic representation [N_x], which would spread its activation to representations containing it, including [$_{NP}$ D N_x]. The syntactic representation of the determiner could be coindexed with the orthographic and/or phonological form of *a,* so that *a gidari* is processed as an NP ("the pilot"). *A gidari* will most likely be processed as an NP than as a PP, for in Spanish it is more frequent to encounter an NP in preverbal position than a PP. For instance, of all sentences with an animate argument in this position in the ADESSE corpus, in just 3.3% this argument is preceded by the preposition *a.* In the remaining 96.70%, this argument is an NP with or without a determiner (the corpus does not allow to further narrow down the search). As usual, this NP will spread its activation to larger (L1) syntactic representations containing it, possibly including [$_{CP}$ [$_{NP}$ D N] [$_{VP}$ V]], [$_{CP}$ [$_{NP}$ D N] [$_{VP}$ V [$_{PP}$ PREP NP]]] and [$_{CP}$ [$_{NP}$ D N] [$_{VP}$ V NP]]. When the (cognate or non-cognate) verb is processed, its activation will further raise the current activation level of the aforementioned structures. These structures contain a finite verb, but the L2 verb is in its citation form. Like for SVO sentences, I propose that learners will simply assume that finite and non-finite verbs have the same form in the L2.

Finally, *antzezle* will be processed. The orthographic and/or phonological representations of this word will activate the syntactic representation of a noun. This will feed back activation to two L1 representations in the syntactic store, [$_{CP}$ [$_{NP}$ D N] [$_{VP}$ V [$_{PP}$ PREP NP]]] and [$_{CP}$ [$_{NP}$ D N] [$_{VP}$ V NP]]. The first structure is more frequent in Spanish than the second one, so it will probably be used to process the input (of all transitive SVO sentences with two animate arguments in the ADESSE corpus, in 75% the object is preceded by the preposition *a*). The syntactic representation of the preposition will probably be coindexed with a null orthographic/phonological representation. Importantly, I assume that, due to the agent-first preference, the first NP (*a gidari*) will be incorrectly interpreted as the agent of the verb (*pintatu/margotu*) and the second NP (*antzezle*), will be incorrectly interpreted as the patient. Specifically, the sentence will be interpreted as "the pilot is painting the actor" but the non-linguistic context accompanying it (in Experiment 5, a picture representing the meaning of the sentence, see Section 5.4) will indicate that it should be interpreted as "the actor is painting the pilot". After misprocessing one or more OVS sentences, learners will realize that the language has flexible word order and a reanalysis from a subject/agent-first structure to an object/patient-first structure will occur. In subsequent processing, an L1 structure such as [$_{CP}$ [$_{PP}$ PREP NP] [$_{VP}$ (CL) V NP]], with the syntactic representation of the preposition coindexed with the orthographic/phonological representations of *a*, will activate and will be used to process the input. This time, the noun preceded by *a* will be correctly interpreted as the patient and the noun not preceded by *a,* as the agent. Each time that OVS sentences with prepositional patient marking are processed in this way, the resting activation level of this structure will increase.

I discuss now the role of cognates in sentence processing. First, cognate verbs might be processed faster than non-cognate verbs due to their higher current and resting activation level (see Section 1.1), causing overall sentence processing to be faster when this includes a cognate than a non-cognate. As mentioned, I assume that OVS sentences will be initially misanalysed as SVO. If sentences are processed faster when they contain a cognate verb than a non-cognate verb, then it could be that the mismatch between the interpretation of the sentence and the picture accompanying it is detected faster for sentences with cognates than for sentences with non-cognates. However, SVO is the preferred word order in Spanish and processing SVO sentences should be naturally very easy. Consequently, it could be that, overall, cognates do not particularly facilitate that learners detect the mismatch between their interpretation of an OVS sentence and the picture, what eventually triggers the reanalysis of the sentence from an SVO to an OVS order. Once this type of sentences are reanalysed in terms of an object-first structure, cognates may affect processing in two different ways, depending on whether the L1 object-first structure, stored in learners' linguistic system at the beginning of L2 acquisition, has a low or a high resting activation level. If the structure is stored *with a low resting activation level* (for it is not very frequent in the L1), then the cognate effect could be similar to the one for the cross-linguistically dissimilar L2 structures in Experiment 4. That is, the current activation level of the structure should be higher when processed with a cognate verb than with a non-cognate verb due to a stronger activation spreading from the syntactic representation of the cognate than the non-cognate to the structure. This should translate into a higher resting activation level for the structure processed with a cognate (12a) than for the one processed with a non-cognate (12b) and, thus, into the former being more robustly established in learners' linguistic system than the latter. On the other hand, if the object-first structure is stored in the linguistic system *with a high resting activation level* (because, even if not very frequent, it has been processed multiple times in the L1), then it is possible that the situation described does not hold. That is, because the increases in resting activation level after processing diminish as experience with a structure increases, processing L2 sentences with this structure should cause small increases in its resting activation level. As a result, the difference in resting level derived from processing the structure with cognate vs. non-cognate verbs may be non-significant.

In any case, the SVO structure with prepositional patient marking should be more robustly established in (cognate and non-cognate) learners' linguistic system than the OVS structure with prepositional patient marking. This is because in Spanish, the former is much more frequent than the latter (e.g. of all transitive sentences in which the direct object is preceded by the preposition *a* in the ADESSE corpus, 92.6% have an SVO order and just 7.4% have an OVS order). I assume that the SVO structure will be comparably established in cognate and non-cognate learners' linguistic system. If the OVS structure is more robustly established in cognate learners' than in non-cognate learners' linguistic system, then the difference between the SVO structure and the OVS structure should be smaller when processed with cognates than with non-cognates. By contrast, if the OVS structure is comparably established in cognate and non-cognate learners' linguistic system, then the difference between the SVO and the OVS structure should be similar when processed with cognates and with non-cognates.

5. Experiment 5

5.1. Overview

The research question of Experiment 5 was "*Do cognates facilitate the initial acquisition of cross-linguistically similar L2 structures?"*. To answer this question, I exposed two groups of Spanish natives without knowledge of Basque to the Spanish-based structures with Basque non-cognate nouns and a Spanish-Basque cognate or non-cognate verb exemplified in Section 4.2. Then, I assessed whether exposure to the structures with cognates, as opposed to non-cognates, affected the establishment of these structures in the linguistic system as hypothesized in that section. Experiment 5 was identical to Experiment 4, except for the structures to be learnt. Therefore, the cognate and non-cognate versions of the mini-language used in Experiment 4 were also used in Experiment 5. The two versions included Basque non-cognate nouns and Spanish-Basque either cognate or non-cognate verbs. These were used to create sentences exemplifying the target structures. As in Experiment 4, I will use the terms *cognate learners* and *non-cognate learners* to refer to participants who learnt the structures with cognate verbs and with non-cognate verbs, respectively. The procedure of Experiment 5 was like that of Experiment 4. First, cognate and non-cognate learners learnt the non-cognate nouns and either the cognate or the non-cognate verbs by means of picture-word association. Next, there was an exposure phase, during which the two groups of learners read and listened to SVO and OVS sentences with prepositional patient marking and with a cognate or a non-cognate verb, each accompanied by a picture. Then, a second vocabulary-learning phase followed, during which cognate and non-cognate learners learnt novel non-cognate verbs. These were used in the testing phase, which consisted of a sentence-picture congruency task and a written production task. Finally, a debriefing phase encouraged learners to report their syntax knowledge. I postulated two possible hypotheses, each divided into two parts. These were advanced in Section 4.2 and can be summarised as follows.

Hypothesis 1a (H1a) claimed that the target structures, with verb-medial word order and prepositional patient marking, would be comparably established in the linguistic system as a result of processing sentences with cognate verbs and with non-cognate verbs. In familiar thinking terms, I hypothesized that cognates would not facilitate the acquisition of the L1-L2 similar structures. If this hypothesis was retained, then I proposed **Hypothesis 1b (H1b)**. This claimed that the SVO structure would be more robustly established in learners' linguistic system than the OVS structure and that the difference between the two structures would be comparable when processed with cognate verbs and with non-cognate verbs. That is, I hypothesized that there would be a learning advantage for the SVO structure over the OVS structure and that this advantage would be comparable for cognate and non-cognate learners.

Hypothesis 2a (H2a) claimed that the SVO structure with prepositional patient marking would be comparably established in the linguistic system as a result of processing sentences with cognate verbs and with non-cognate verbs. By contrast, the OVS structure with prepositional patient marking would be more robustly established in the linguistic system as a result of processing sentences with cognate verbs than with non-cognate verbs. In familiar thinking terms, I hypothesized that cognates would facilitate the acquisition of the OVS structure, but not of the SVO structure. If this hypothesis was retained, then I proposed **Hypothesis 2b (H2b)**. This claimed that the SVO structure would be more robustly established in learners' linguistic system than the OVS structure, but that the difference between the two structures would be smaller when processed with cognate verbs than with non-cognate verbs. That is, I hypothesized that there would be a learning advantage for the SVO structure over the OVS one, but that this advantage would be smaller for cognate learners than for non-cognate learners.

5.2. Participants

Participants were 60 Spanish natives (53 female) aged 18–33 (*M* = 21, *SD* = 2.85) studying at the University of Barcelona (UB). They were divided into two groups of 30 cognate learners and 30 non-cognate learners. Cognate and non-cognate learners filled out the same linguistic background questionnaire as in Experiment 4 (Appendix C-1). All participants indicated having no knowledge of Basque. Since most young Spanish natives in Catalonia are bilingual with Catalan, the questionnaire asked about language history, proficiency and use in the two languages. Most participants (90% of cognate learners; 90% of non-cognate learners) reported at least some knowledge of Catalan[19]. Yet, they all indicated feeling more comfortable using Spanish and speaking in Spanish with one or both of their parents before starting school (0-3 years). Cognate and non-cognate learners rated their frequency of use of Spanish and Catalan on a 7-point scale (1 = *Spanish only*, 7 = *Catalan only*) in their childhood (3-12 years), puberty (12-18 years) and adulthood (after 18 years) at school/university/work, at home and at other places. The mean scores for cognate and non-cognate learners in each period and environment can be consulted in Appendix C-2. In short, the two groups of learners were Spanish-dominant, since Spanish was the language they used and had used the most throughout their lives (mean language use during childhood, cognate learners, 2.89 (*SD* = 1.67), non-cognate learners, 3.07 (*SD* = 1.96); puberty, cognate learners, 2.62 (*SD* = 1.43), non-cognate learners, 2.88 (*SD* = 1.71); adulthood, cognate learners, 2.61 (*SD* = 1.21), non-cognate learners, 2.69 (*SD* = 1.31)). There were no significant differences in the mean frequency of use of Spanish between groups, as revealed by independent-samples t-tests (all $p \geq .28$, see Appendix C-2). The two groups also self-rated their proficiency speaking, listening, reading and writing in Spanish on a 7-point scale (1 = *very poor proficiency*, 7 = *perfect proficiency*). Cognate and non-cognate learners' mean scores for the four skills are also reported in Appendix C-2. The two groups reported having a nearly perfect and comparable mastery of Spanish (mean proficiency collapsing all skills, cognate learners, 6.82 (*SD* = 0.45) vs. non-cognate learners, 6.86 (*SD* = 0.37), t (238) = -0.78, p = .44, d = -0.10). All participants indicated having normal or

[19] The fact that participants knew Catalan was not judged problematic, since in Catalan SVO and OVS word orders are allowed and it is possible to use the preposition *a* to mark the patient of the transitive sentence, as in Spanish. From the point of view of normative grammar, *a* should not be used. Yet, it is widespread in most varieties (Pineda, 2021).

corrected to normal vision and hearing. Prior to the experiment, they read and signed an informed consent (Appendix C-3). Experiment 5 was part of the project “Cross-linguistic activation effects in bilingual language processing and learning” (PGC2018-097970-B-I00), funded by the Spanish Ministry of Science, Innovation and Universities and approved by the Committee of Ethics for research involving human beings of the University of the Basque Country (*Comité de Ética para las Investigaciones con Seres Humanos, CEISH,* Ref. M10_2019_167). Participation was rewarded with 12€.

5.3. Materials

The cognate and non-cognate versions of the mini-language in Experiment 4 were used in Experiment 5, with the exception of the structures. The vocabulary was identical to that in Experiment 4. For the cognate version of the language, this consisted of five Basque nouns of professions, four Spanish-Basque cognate verbs used in the exposure phase and four Spanish-Basque non-cognate verbs used in the testing phase. The non-cognate version of the language included the same nouns and non-cognate verbs as the cognate version, but cognate verbs were replaced by their non-cognate synonyms. The word-picture pairs created for Experiment 4 (available in Appendix C-4) were also used in Experiment 5, in the vocabulary-learning phases preceding the exposure phase and the testing phase. These pairs were used to create sentence-picture pairs for the exposure phase and the testing phase (the exposure set and the testing set). The audios of the individual nouns and verbs used in Experiment 4, recorded by a male native speaker of Basque, were also used in Experiment 5. In addition, this speaker recorded the sentences in the exposure and the testing sets. Sentences were recorded at a normal pace and with natural intonation in a soundproof booth using an Olympus voice recorder (Linear PCM Recorder LS-5 model, frequency sampling of 96kHz). Recordings were cut using *Praat* (Boersma & Weenink, 2018, version 6.0.37). In the next sections, I describe the exposure set and the testing set.

5.3.1. Exposure set

The exposure set in Experiment 5 was derived from the exposure set in Experiment 4. In that experiment, this set consisted of 80 baseline SOV sentences with postpositional agent-patient marking and a cognate verb, from which I created three other versions varying word order (SOV vs. OSV) and the cognate status of the verb (cognate vs. non-cognate) (cf. Section 3.3.1). In Experiment 5, I transformed these SOV and OSV sentences with postpositional agent-patient marking into SVO and OVS sentences with prepositional patient marking (see (13) for an example with a Spanish-Basque cognate verb, all sentences meaning “The actor is painting the pilot”).

(13)

		Experiment 4				**Experiment 5**	
	S	O	V	→	S	V	O
a.	antzezle-ak	gidari-a	pintatu		antzezle	pintatu	a gidari
	actor-A	pilot-P	paint		actor.A	paint	P pilot
	O	S	V	→	O	V	S
b.	gidari-a	antzezle-ak	pintatu		a gidari	pintatu	antzezle
	pilot-P	actor-A	paint		P pilot	paint	actor.A

The sentence-picture pairs in Experiment 4 were also used in Experiment 5. The 160 sentence-picture pairs with a cognate verb (80 SVO and their 80 OVS counterparts) were assigned to the cognate version of the language and the 160 sentence-picture pairs with a non-cognate verb (80 SVO and their 80 OVS counterparts) were assigned to the non-cognate version of the language. The sentence-picture pairs for the cognate and the non-cognate versions were distributed into two lists, so that participants would not hear and see an SVO sentence and its OVS equivalent. Therefore, each participant was exposed to 80 sentence-picture pairs: 40 SVO and 40 OVS. Table 4.11 exemplifies an SVO sentence-picture pair and the corresponding OVS sentence-picture pair for each version of the language.

	Language version	Sentence	Picture
SVO	Cognate	Antzezle pintatu a gidari actor.A paint P pilot "The actor is painting the pilot."	
	Non-cognate	Antzezle margotu a gidari actor.A paint P pilot "The actor is painting the pilot."	
OVS	Cognate	A gidari pintatu antzezle P pilot paint actor.A "The actor is painting the pilot."	
	Non-cognate	A gidari margotu antzezle P pilot paint actor.A "The actor is painting the pilot."	

TABLE 4.11. Examples of SVO and OVS sentence-picture pairs for the cognate and the non-cognate version of the language in Experiment 5. The four sentences have the same meaning and, thus, they are paired with the same picture.

The lexicon of the exposure set consisted of the same five non-cognate nouns and four cognate/non-cognate pairs of verbs as in Experiment 4. As a reminder, the cognate verbs and their Spanish translations had identical or nearly identical stems (orthographic LD, *M* = 0.25, *SD* = 0.43; phonological LD, *M* = 0.5, *SD* = 0.5) and different endings. Cognate verbs and their non-cognate synonyms were matched in length (in number of letters). Each noun appeared twice as the agent and twice as the patient of each verb in SVO and OVS sentences per list. The combination of a given agent, patient and verb occurred only once in the same order in the exposure set. This can be consulted in Appendix C-8.

5.3.2. Testing set

5.3.2.1. Sentence-picture congruency task

The testing phase started with a sentence-picture congruency task. The testing set in Experiment 5 was derived from the testing set in Experiment 4. For that experiment, I created 80 SOV sentences with postpositional agent-patient marking and with a non-cognate verb as well as 80 OSV sentences derived from the SOV ones. Each pair of SOV-OSV sentences was associated with four different pictures to generate syntactically congruent and incongruent sentence-picture pairs (testing learning of the target structures) and semantically incongruent sentence-picture pairs with agent or patient violation (distractors testing vocabulary knowledge). In sum, there were eight sentence-picture pairs derived from an SOV sentence: (i) a syntactically congruent SOV sentence-picture pair, (ii) a syntactically incongruent SOV sentence-picture pair, (iii) a semantically incongruent SOV sentence-picture pair with agent violation, (iv) a semantically incongruent SOV sentence-picture pair with patient violation and the four equivalent OSV sentence-picture pairs. These pairs were shared for the cognate and the non-cognate language versions (cf. Section 3.3.2).

To create the testing set in Experiment 5, I transformed the SOV and OSV sentences with postpositional agent-patient marking into SVO and OVS sentences with prepositional patient marking, as exemplified in (13) for the exposure set. Apart from this change, the sentence-picture pairs in Experiments 4 and 5 were identical, and so was their distribution into eight lists of 80 sentence-picture pairs (40 SVO and 40 OVS: 10 syntactically congruent, 10 syntactically incongruent, 10 semantically incongruent with agent violation and 10 semantically incongruent with patient violation per word order). A list never included an SVO sentence and its OVS equivalent, it never included a syntactically congruent sentence-picture pair and its incongruent counterpart nor a semantically incongruent sentence-picture pair with agent violation and its equivalent sentence-picture pair with patient violation. A given SVO or OVS sentence appeared only once per list and a given picture only occurred once within condition per list. Table 4.12 exemplifies the four types of sentence-picture pairs for an SVO sentence and its OVS equivalent.

	Pair type	Sentence	Picture
SVO	Syntactically congruent	Antzezle aukeratu a gidari actor.A choose P pilot "The actor is choosing the pilot."	
	Syntactically incongruent	Antzezle aukeratu a gidari actor.A choose P pilot "The actor is choosing the pilot."	
	Semantically incongruent with agent violation	Antzezle aukeratu a gidari actor.A choose P pilot "The actor is choosing the pilot."	
	Semantically incongruent with patient violation	Antzezle aukeratu a gidari actor.A choose P pilot "The actor is choosing the pilot."	
OVS	Syntactically congruent	A gidari aukeratu antzezle P pilot choose actor.A "The actor is choosing the pilot."	
	Syntactically incongruent	A gidari aukeratu antzezle P pilot choose actor.A "The actor is choosing the pilot."	
	Semantically incongruent with agent violation	A gidari aukeratu antzezle P pilot choose actor.A "The actor is choosing the pilot."	
	Semantically incongruent with patient violation	A gidari aukeratu antzezle P pilot choose actor.A "The actor is choosing the pilot."	

TABLE 4.12. Examples of the four types of sentence-picture pairs generated from an SVO sentence and its OVS equivalent in Experiment 5. The sentence-picture pairs were shared for the cognate and the non-cognate version of the language.

The lexicon of the testing set was identical to the one in Experiment 4: five non-cognate nouns and four non-cognate verbs. Each noun appeared four times as the agent and four times as the patient of each verb per list: once in an SVO sentence-picture pair testing syntax learning, once in an OVS sentence-picture pair testing syntax learning, once in an SVO sentence-picture pair testing vocabulary learning and once in an OVS sentence-picture pair testing vocabulary learning. Whether these were syntactically congruent or incongruent sentence-picture pairs and semantically incongruent pairs with agent or patient violation was counterbalanced across lists. The full set of sentence-picture pairs can be found in Appendix C-8.

5.3.2.2. Production task

The testing phase also included a written picture-description task. The pictures for this task (Appendix C-8) were the same as in Experiment 4. These were eight pictures selected from eight sentence-picture pairs within the ones in the sentence-picture congruency task and were shared for the cognate and the non-cognate language versions. Each of the five possible nouns acted as agent and as patient in at least one picture. Each of the four possible verbs appeared in two pictures. Two nouns appeared together in only one picture.

5.4. Procedure

Participants were informed that they would learn some words in Basque and then perform some comprehension tasks. They were not told that they participated in a syntax learning experiment. Experiment 5 consisted of five phases: a first vocabulary-learning phase, an exposure phase, a second vocabulary-learning phase, a testing phase and a debriefing phase in the form of a verbal report. Cognate and non-cognate learners also conducted a reading span task, which controlled for group differences in working memory capacity. The procedure of the experiment was like in Experiment 4 (Sections 3.4.1 to 3.4.6). This procedure is summarised below:

- **First vocabulary-learning phase:** Cognate and non-cognate learners learnt the same five non-cognate nouns and either four cognate verbs or their non-cognate synonyms. In each trial, participants saw a picture, the Basque noun or verb it represented written below and the Spanish translation between brackets. The Basque word was simultaneously played. Participants repeated the word aloud and then pressed the space bar to move on to the next trial. Each word-picture pair was shown four times. The order of the pairs was pseudo-randomized; nouns were presented before verbs. Learning was tested in a picture-word matching task and in a picture-naming task, both performed until 100% accuracy.
- **Exposure phase:** Cognate and non-cognate learners saw pictures of transitive actions involving two nouns and a verb out of the learnt ones, each visually and aurally accompanied by a sentence describing it. The two groups were exposed to SVO and OVS sentences with prepositional patient marking and with cognate or non-cognate verbs, respectively. They were instructed to look at each picture and listen to and read the

accompanying sentence. Sentence-picture pairs were randomized and they were automatically presented one after the other.

- **Second vocabulary-learning phase:** Cognate and non-cognate learners learnt four new non-cognate verbs, which would later be used in the testing phase. These verbs were learnt and tested as in the first vocabulary-learning phase.
- **Testing phase:** The testing phase consisted of a sentence-picture congruency task and a picture-description task. In the first task, cognate and non-cognate learners judged four types of SVO and OVS sentence-picture pairs: (i) syntactically congruent, (ii) syntactically incongruent, (iii) semantically incongruent with agent violation and (iv) semantically incongruent with patient violation. In each trial, learners saw a picture accompanied by a sentence in written and oral form. When the audio of the sentence finished, the options *Correcto* ("Correct") and *Incorrecto* ("Incorrect") appeared below the sentence-picture pair. Learners had to judge, as quickly as possible, whether the sentence was a correct description of the picture or not by pressing the key "A" or the key "L", respectively. They had 5 seconds to respond and received no feedback on the accuracy of their responses. In the second task, cognate and non-cognate learners saw pictures of transitive actions and a list of all the nouns and verbs learnt. They had to write a sentence describing each picture choosing the appropriate words from the list. There was no time limit to write the sentences.
- **Debriefing phase:** Learners were asked some questions to elicit knowledge about the patient marking in the language, the part of the experiment in which they became aware of this marking and any strategies used in the sentence-picture congruency task (e.g. intuition, syntax knowledge or other). These questions were:
    1. In all the sentences you have heard there were two nouns, a verb and another word. Could you say what this other word was?
    2. Could you say why it appeared in the sentence or how it was used?
    3. In which part of the experiment did you realize this?
    4. Did you follow any strategy to perform the test?[20]
- **Reading span task:** Cognate and non-cognate learners conducted the Spanish version of Unsworth et al.'s (2005) reading span task, which requires remembering individual letters while performing plausibility judgements.

All tasks were run on the E-prime 3.0 software. Participants were tested individually in a soundproof booth and listened to audio files through headphones. The experiment took around 90 minutes to complete. Instructions were administered in Spanish and were identical to the ones in Experiment 4, with the exception of the questions in the debriefing phase (see Appendix C-6).

[20] I read the questions with the participants and made sure that they understood that the "test" in the fourth question referred to the sentence-picture congruency task. I also elicited any strategies followed, when necessary.

5.5. Predictions

In Section 5.1, I proposed two hypotheses about how being exposed to the two cross-linguistically similar L2 structures with cognate verbs or non-cognate verbs would affect the initial establishment of these structures in learners' linguistic system. Experiment 5 aimed to test these hypotheses, which addressed the research question of the study ("*Do cognates facilitate the initial acquisition of cross-linguistically similar L2 structures?*"). In this section, I will review these hypotheses and discuss their predictions. In addition, like in Experiment 4, in Experiment 5 cognate and non-cognate learners learnt non-cognate nouns, cognate or non-cognate verbs in a first vocabulary-learning phase and novel non-cognate verbs in a second vocabulary-learning phase. Hence, some predictions could also be made regarding this lexical learning. In what follows, I present my predictions for vocabulary learning and for syntax learning.

Predictions for vocabulary learning

My predictions for vocabulary learning were the same as in Experiment 4 (Section 3.5) and are summarised here. In the first and second vocabulary-learning phases, cognate and non-cognate learners learnt nouns and verbs until obtaining 100% accuracy in the picture-word matching task and in the picture-naming task. First, I predicted that, since in the first vocabulary-learning phase the two groups learnt the same nouns, they would need a comparable number of attempts to match all nouns to pictures and to name all pictures of nouns correctly. Additionally, learning of nouns was assessed in the sentence-picture congruency task (testing phase), when learners judged semantically incongruent sentence-picture pairs with agent or patient violation. I predicted that overall accuracy in these sentence-picture pairs would be comparable for the two groups of learners. I did not predict a significantly more accurate performance when judging sentence-picture pairs with agent vs. patient violation neither for cognate nor for non-cognate learners. On the other hand, in the first vocabulary-learning phase cognate and non-cognate learners also learnt cognate or non-cognate verbs, respectively. I predicted that, given the similarity of cognates in the L1 and the L2, cognate verbs would be learnt more easily than non-cognate verbs. In other words, I predicted that cognate learners would need significantly fewer attempts to match all verbs to pictures and to name all pictures of verbs correctly than non-cognate learners. Finally, in the second vocabulary-learning phase cognate and non-cognate learners learnt the same non-cognate verbs. I predicted that the two groups would achieve 100% accuracy in the picture-word matching task and in the picture-naming task in a statistically similar number of attempts.

Predictions for syntax learning

I postulated two hypotheses about the influence that cognates could have on the initial acquisition of the cross-linguistically similar L2 structures. My first hypothesis had two parts. **Hypothesis 1a (H1a)** claimed that the target structures, with verb-medial word order and prepositional patient marking, would be comparably established in the linguistic system as a result of processing sentences with cognate verbs and with non-cognate verbs. That is, cognates would not facilitate the acquisition of the cross-linguistically similar structures. If this hypothesis was retained, then **Hypothesis 1b (H1b)** claimed that there would be a learning advantage for the SVO structure over the OVS one and that this advantage would be comparable for cognate learners and non-cognate learners. My second hypothesis also had two parts. **Hypothesis 2a (H2a)** claimed that the SVO structure with prepositional patient marking would be comparably established in the linguistic system after processing sentences with cognate verbs and with non-cognate verbs. Conversely, the OVS structure with prepositional patient marking would be more robustly established as a result of processing sentences with cognates than with non-cognates. In other words, cognates would facilitate the acquisition of the OVS structure, but not of the SVO structure. If this hypothesis was correct, then **Hypothesis 2b (H2b)** claimed that there would be a learning advantage for the SVO structure over the OVS one, but that this would be smaller for cognate learners than for non-cognate learners. A prerequisite for one of these two hypotheses to be met was that the two target structures were part of learners' linguistic system. I made different predictions regarding how this could be seen in the sentence-picture congruency task and in the production task.

In the congruency task, sentences in syntactically congruent sentence-picture pairs were acceptable in the L2, for patient marking was used correctly, i.e. the noun marked as the patient in the sentence was the patient in the picture. By contrast, sentences in syntactically incongruent sentence-picture pairs could be considered unacceptable in the L2, for patient marking was used incorrectly, i.e. the noun marked as the patient in the sentence was the agent in the picture. If the target structures were established in learners' linguistic system, I predicted that cognate and non-cognate learners would distinguish between syntactically congruent and syntactically incongruent SVO and OVS sentence-picture pairs. Specifically, the two groups' mean accuracy when judging the four types of sentence-picture pairs would be significantly above chance. Turning to the production task, if the SVO structure with prepositional patient marking was part of learners' linguistic system, I predicted that cognate and non-cognate learners would correctly write significantly more than 50% of their verb-medial picture descriptions with the preposition *a* preceding the second noun in the sentence. Likewise, if the OVS structure with prepositional patient marking was part of learners' linguistic system, I predicted that they would correctly write significantly more than 50% of their verb-medial picture descriptions with *a* preceding the first noun in the sentence.

If these predictions were met, my two hypotheses will be tested. As mentioned, **H1a** claimed that overall the target structures would be comparably established in cognate and non-cognate learners' linguistic system. **If H1a was correct,** I predicted that, in the congruency task, cognate and non-cognate learners would be comparably accurate in their judgement of syntactically congruent and incongruent sentence-picture pairs. In addition, **H1b** claimed that there would be a learning advantage for the SVO structure over the OVS one, which would be comparable for cognate and non-cognate learners. **If H1b was correct,** I predicted that accuracy in syntactically congruent and incongruent SVO sentence-picture pairs would be significantly higher than in syntactically congruent and incongruent OVS sentence-picture pairs and that this would not significantly vary as a function of group of learners. Additionally or alternatively, the production task would yield similar results. That is, **if H1a was true**, I predicted that overall cognate and non-cognate learners would be comparably accurate in their verb-medial picture descriptions with *a* before one of the two nouns in the sentence. **If H1b was true,** learners would be significantly more accurate when writing verb-medial picture descriptions with *a* before the second noun in the sentence than before the first noun in the sentence, and this difference would be statistically similar for both groups of learners.

On the other hand, **H2a** claimed that cognates would facilitate the acquisition of the OVS structure with prepositional patient marking, but not of the SVO one. **If H2a was correct,** I predicted that, in the congruency task, cognate and non-cognate learners would be comparably accurate in their judgement of syntactically congruent and incongruent SVO sentence-picture pairs, but that the first group of learners would be significantly more accurate than the second one when judging syntactically congruent and incongruent OVS sentence-picture pairs. Additionally, **H2b** claimed that there would be a learning advantage for the SVO structure over the OVS one and that this would be smaller for cognate learners than for non-cognate learners. **If H2b was correct,** I predicted that the two groups of learners would be significantly more accurate when judging SVO sentence-picture pairs than OVS sentence-picture pairs. Yet, this difference would be smaller for cognate learners than for non-cognate learners, as indicated, for instance, by smaller effect sizes and/or larger p-values in the tests comparing accuracy in SVO and OVS sentence-picture pairs. Turning to the production task, **If H2a was true,** I predicted that cognate and non-cognate learners would be equally accurate in their verb-medial picture descriptions with *a* before the second noun in the sentence. However, the former would be significantly more accurate than the latter in their picture descriptions with *a* before the first noun in the sentence. **If H2b was true,** I predicted that learners would be significantly more accurate when writing verb-medial picture descriptions with *a* before the second noun in the sentence than before the first noun in the sentence. Yet, this difference would be smaller for cognate learners than for non-cognate learners (smaller effect sizes and/or larger p-values in the tests comparing accuracy in the two types of picture descriptions).

5.6. Coding and data analysis

I analysed Experiment 5 using the programming environment R (R Core Team, 2022, version 4.2.2). The function and package with which each statistical test and effect size measure were calculated are the same as reported for Experiment 4.

5.6.1. First vocabulary-learning phase

5.6.1.1. Picture-word matching task

Accuracy was measured after the selection of a word in each trial (1 = correct picture-word matching, 0 = incorrect picture-word matching). Cognate and non-cognate learners' number of attempts at picture-noun matching and picture-verb matching were coded as positive integers. The two groups' number of attempts at picture-noun matching were not normally distributed, as indicated by Shapiro-Wilk tests (cognate learners, $W = 0.40$, $p < .001$; non-cognate learners, $W = 0.40$, $p < .001$). Cognate and non-cognate learners attempted picture-noun matching the same number of times, so no further analyses were conducted on the data. The two groups' number of attempts at picture-verb matching were not normally distributed either (cognate learners, $W = 0.55$, $p < .001$; non-cognate learners, $W = 0.67$, $p < .001$). A Wilcoxon rank-sum test with continuity correction compared the distributions of the number of attempts of the two groups of learners to examine whether these had similar shapes, i.e. whether one of the two distributions had significantly larger values than the other one. The standardised measure of effect size r was calculated for this test throughout the experiment. An r of 0.1-0.3, 0.3-0.5 and ≥ 0.5 was considered small, medium and large, respectively.

5.6.1.2. Picture-naming task

I coded the number of attempts it took each cognate and non-cognate learner to name all pictures of nouns and verbs correctly as positive integers. Shapiro-Wilk tests revealed that the number of attempts at naming pictures of nouns were not normally distributed neither for cognate learners ($W = 0.72$, $p < .001$) nor for non-cognate learners ($W = 0.78$, $p < .001$). Likewise, the two groups' number of attempts at naming pictures of verbs were non-normally distributed (cognate learners, $W = 0.40$, $p < .001$; non-cognate learners, $W = 0.78$, $p < .001$). Wilcoxon rank-sum tests with continuity correction compared cognate and non-cognate learners' number of attempts at noun picture naming and at verb picture naming.

5.6.2. Second vocabulary-learning phase

5.6.2.1. Picture-word matching task

Accuracy was measured following the selection of a word in each trial (1 = correct picture-word matching, 0 = incorrect picture-word matching). Participants' number of attempts at the task were coded as positive integers. The number of attempts by cognate and non-cognate learners did not follow a normal distribution (cognate learners, $W = 0.51$, $p < .001$; non-cognate learners, $W = 0.56$, $p < .001$). The distributions of the two groups' number of attempts at picture-word matching were compared with a Wilcoxon rank-sum test with continuity correction.

5.6.2.2. Picture-naming task

Cognate and non-cognate learners' number of attempts at picture naming were coded as positive integers. The two groups' number of attempts were not normally distributed (cognate learners, $W = 0.49$, $p < .001$; non-cognate learners, $W = 0.59$, $p < .001$). The distributions of the number of attempts by cognate and non-cognate learners were compared with a Wilcoxon rank-sum test with continuity correction.

5.6.3. Testing phase

5.6.3.1. Sentence-picture congruency task

In this section, I present how I coded and analysed the output of the sentence-picture congruency task. This comprises the strategies that participants used to perform the task, even if such strategies were elicited in the debriefing phase.

<u>Performance on semantically incongruent sentence-picture pairs</u>

I measured accuracy after judging each sentence-picture pair (1 = correct congruency judgement, 0 = incorrect congruency judgement). I eliminated trials in which participants did not respond: 3.67% (44/1200) of all sentence-picture pairs seen by cognate learners and 3.17% (38/1200) of all pairs seen by non-cognate learners. As a reminder, all semantically incongruent sentence-picture pairs contained an agent or a patient violation, i.e. the agent or the patient in the sentence did not match the ones in the picture. I tested whether cognate and non-cognate learners were equally accurate when judging these two types of sentence-picture pairs, which would corroborate that knowledge of nouns was comparable between groups. A generalized linear mixed effects model looked into the interaction between the effect of Group of learners (Cognate vs. Non-cognate) and Type of sentence-picture pair (with Agent violation vs. with Patient violation) on accuracy. The model had by-participant and by-item random intercepts and a by-participant random slope of Type of sentence-picture pair. Deviation coding was used for the variables Group of learners (*Cognate* coded as 0.5 and *Non-cognate*, as -0.5) and Type of sentence-picture pair (*with Agent violation* coded as 0.5 and *with Patient violation*, as -0.5).

<u>Performance on syntactically congruent and incongruent sentence-picture pairs</u>

Accuracy was measured after judging each sentence-picture pair (1 = correct congruency judgement, 0 = incorrect congruency judgement). Trials in which participants did not respond were removed. These were 4.67% (56/1200) of all sentence-picture pairs seen by cognate learners and 4.92% (59/1200) of all pairs seen by non-cognate learners. More precisely, I eliminated 2.67% (8/300) of all syntactically congruent SVO sentence-picture pairs, 4.67% (14/300) of all syntactically congruent OVS sentence-picture pairs, 5.33% (16/300) of all syntactically incongruent SVO sentence-picture pairs and 6% (18/300) of all syntactically incongruent OVS sentence-picture pairs seen by cognate learners. Likewise, I eliminated 3.33% (10/300) of all syntactically congruent SVO sentence-picture pairs, 4% (12/300) of all syntactically congruent OVS sentence-picture pairs, 5.33% (16/300) of all syntactically incongruent SVO sentence-picture pairs and 7% (21/300) of all syntactically incongruent OVS sentence-picture pairs seen by non-cognate learners.

The structures as part of the linguistic system

I first examined whether the two target structures, with verb-medial word order and prepositional patient marking, were part of cognate and non-cognate learners' linguistic system. I calculated the two groups' mean accuracy percentages when judging syntactically congruent and incongruent SVO and OVS sentence-picture pairs. One-sample t-tests compared each percentage against chance. As will be detailed in Section 5.7.3, the analysis indicated that the SVO structure was established in learners' linguistic system, but there was no evidence that this was true for the OVS structure. Considering this, I tested whether cognate and non-cognate learners' accuracy in SVO sentence-picture pairs (collapsing congruent and incongruent pairs) was comparable. I fitted a generalized linear mixed effects model that had Accuracy as a dependent variable, Group of learners as independent variable and random intercepts by participant and by item[21]. Treatment coding was used for the variable Group of learners, with the category *Cognate* coded as 0 and the category *Non-cognate,* as 1. Then, I tested whether cognate and non-cognate learners had an overall tendency to judge OVS sentence-picture pairs as congruent or "correct". I coded responses in the congruency task as *Hits* (congruent sentence-picture pair judged as "correct"), *False alarms* (incongruent sentence-picture pair judged as "correct"), *Misses* (congruent sentence-picture pair judged as "incorrect") or *Correct rejections* (incongruent sentence-picture pair judged as "incorrect"). Next, I calculated the mean index of response bias *c* for syntactically congruent and incongruent OVS sentence-picture pairs for the two groups of learners. One-sample t-tests compared these indices against zero. The analysis revealed a significant response bias indicating that learners tended to judge syntactically congruent and incongruent OVS sentence-picture pairs as congruent.

To have a better understanding of the data, I also examined whether cognate and non-cognate learners' accuracy differed when judging congruent and incongruent OVS sentence-picture pairs. A generalized linear mixed effects model tested for the interaction between the effect of Group of learners (Cognate vs. Non-cognate) and Congruency of the sentence-picture pair (Congruent vs. Incongruent) on accuracy. The model that fitted the data best included random intercepts by participant and by item[22]. I used deviation coding for the two independent variables. The categories *Cognate* and *Congruent* were coded as 0.5 and the categories *Non-cognate* and *Incongruent,* as -0.5. Since the model yielded a significant interaction (see Section 5.7.3), post-hoc generalized linear mixed effects models tested whether cognate and non-cognate learners' accuracy significantly differed (i) in syntactically congruent sentence-picture pairs and (ii) in syntactically incongruent sentence-picture pairs. The models that converged

[21] A by-item random slope of Group of learners did not provide a better fit for the data, as indicated by nested model comparisons: $X^2(2) = 3.35$, $p = .19$.

[22] Comparison of the models with and without a by-participant random slope of Congruency of the sentence-picture pair: $X^2(2) = 2.68$, $p = .26$.

had Accuracy as dependent variable, Group of learners as independent variable (with *Cognate* coded as 0 and *Non-cognate*, as 1) and random intercepts by participant and by item.

<u>Strategies used to perform the sentence-picture congruency task</u>

Finally, I conducted a descriptive analysis of the strategies that cognate and non-cognate learners reported using in the sentence-picture congruency task (to be consulted in Appendix C-9). As will be detailed in Section 5.7.3, the analysis indicated that the response bias for OVS sentence-picture pairs could be partially the result of a conscious strategy, namely to judge sentence-picture pairs as congruent or incongruent based only on the congruency between the nouns in the sentence and its paired picture. This strategy could have also affected performance on SVO sentence-picture pairs. In this light, sensitivity to the difference between syntactically congruent and incongruent pairs could not be reliably assessed in this task, not even with a *d'* analysis, which separates sensitivity from *unconscious* response bias. Hence, I performed no further analyses.

5.6.3.2. Production task

The structures as part of the linguistic system

As in the sentence-picture congruency task, I started by analysing whether there was evidence that the target structures were part of cognate and non-cognate learners' linguistic system. In the production task, learners wrote sentences to describe pictures of transitive actions using the same nouns and verbs as in the congruency task. First, I calculated the proportion of sentences with SVO and OVS order produced by cognate and non-cognate learners to determine whether the two groups had learnt that there were two word orders in the mini-language. I coded picture descriptions as *SVO* or *OVS* just looking at the order of the nouns and the verb in the sentence. I considered a picture description SVO if the subject/agent in the picture appeared in sentence-initial position and OVS if the object/patient in the picture appeared in sentence-initial position.

Next, I coded whether cognate and non-cognate learners used *a* before a noun in each of their picture descriptions (1 = picture description with *a*, 0 = picture description without *a*). I calculated the percentage of descriptions with and without *a* written by the two groups of learners. One-sample t-tests determined whether the two groups used *a* in significantly more than 50% of sentences to assess whether overall they had learnt that *a* had to be placed before one of the nouns in the sentence. A generalized linear mixed effects model assessed whether cognate and non-cognate learners produced a comparable amount of picture descriptions with *a.* The model that converged had Group of learners (Cognate vs. Non-cognate) as independent variable and random intercepts by participant and by item. Treatment coding was used for the independent variable (*Cognate* coded as 0 and *Non-cognate* coded as 1). Then, I calculated the percentage of picture descriptions in which cognate and non-cognate learners wrote *a* before the first noun in the sentence or before the second noun in the sentence. This was done to determine whether both groups had learnt that *a* could appear in these two positions. Finally, I coded accuracy in the use of *a* as a binary variable (1 = correct use of *a*, 0 = incorrect use of *a*). One-sample t-tests assessed whether cognate and non-cognate learners used *a* correctly in significantly more than 50% of verb-medial sentences with *a* before the second noun and with *a* before the first noun. This would evidence that the SVO and the OVS structures with prepositional patient marking were established in learners' linguistic system. As will be detailed in Section 5.7.3, the tests revealed that this was the case. Consequently, I subsequently examined whether exposure to the structures with cognate verbs vs. non-cognate verbs affected acquisition as predicted by my hypotheses.

Comparing the establishment of the structures in cognate and non-cognate learners' linguistic system

I assessed whether cognate and non-cognate learners' accuracy when writing verb-medial picture descriptions with *a* before the first noun in the sentence and before the second noun in the sentence significantly differed. A generalized linear mixed effects model tested for the interaction between the effect of Group of learners (Cognate vs. Non-cognate) and Position of *a* (Before the first noun vs. Before the second noun) on accuracy. To avoid convergence issues, the model had random intercepts by participant and by item. Deviation coding was used for the variables Group of learners and Position of *a*, with the categories *Cognate* and *Before the first noun* assigned the value 0.5 and the categories *Non-cognate* and *Before the second noun* assigned the value -0.5.

5.6.4. Debriefing phase

I transcribed cognate and non-cognate learners' verbal reports. First, learners were asked whether they knew which word, apart from the nouns and verbs learnt, appeared in the experimental sentences. Then, they were asked to verbalize why it appeared in a sentence or how it was used. I calculated the percentage of cognate and non-cognate learners who (i) reported noticing that sentences included *a* and who (ii) evidenced knowing why or how this word was used. An external researcher and I used this information and a rubric to classify learners as *aware* or *unaware* of patient marking (the transcriptions and the rubric are available in Appendix C-9). Aware learners could state or exemplify that *a* preceded the noun acting as patient/object of the sentence. Alternatively, their answers in the verbal report generally indicated knowledge of patient marking (see Section 5.7.4 for the details). Unaware learners did not report that *a* appeared in the experimental sentences or reported it, but could not (correctly) specify its function. Next, awareness was coded as binary (1 = aware participant, 0 = unaware participant) to calculate the percentage of aware and unaware cognate and non-cognate learners. Finally, I looked at whether aware learners indicated that they realized what the function of *a* was during the exposure phase, which would suggest that learning occurred during that phase and not during the testing phase.

5.6.5. Reading span task

Data from a cognate learner was lost due to equipment malfunction. I collected the partial reading span score for the remaining 59 learners. Cognate and non-cognate learners' scores were normally distributed, as indicated by Shapiro-Wilk tests (W = 0.97, p = .69 and W = 0.94, p = .09, respectively). A Levene's test indicated that the two samples had equal variances (F (1, 57) = 1.17, p = .28). Considering this, I compared cognate and non-cognate learners' mean reading span scores using a two-sample t-test.

5.7. Results

5.7.1. First vocabulary-learning phase

5.7.1.1. Picture-word matching task

The picture-word matching task was performed until cognate and non-cognate learners correctly matched all non-cognate nouns and all cognate or non-cognate verbs with their pictures.

<u>Picture-noun matching</u>

Table 4.13 shows cognate and non-cognate learners' mean and median number of attempts at the task, dispersion measures and 95% confidence intervals for the two values. Twenty-six cognate and non-cognate learners matched all nouns with their pictures in their first attempt at the task. The remaining four cognate and non-cognate learners did so in their second attempt. Since the two groups' performance was identical, their mean and median number of attempts was the same.

	Cognate learners	**Non-cognate learners**
Mean	1.13	1.13
SD	0.35	0.35
95%CI	[1.00, 1.26]	[1.00, 1.26]
Median	1	1
MAD	0	0
95%CI	[1,1]	[1,1]

TABLE 4.13. Information regarding the number of attempts at the picture-noun matching task in Experiment 5 for cognate and non-cognate learners. This includes the mean, Standard Deviation (SD) and the 95% Confidence Interval (95%CI) of the mean, the median, the Median Absolute Deviation (MAD) and the 95%CI of the median.

<u>Picture-verb matching</u>

Cognate and non-cognate learners correctly matched all cognate or non-cognate verbs with their pictures in either one or two attempts at the task, with the exception of a non-cognate learner who did so in a third attempt. More precisely, 73% (22/30) of cognate learners performed the task once, while the remaining 27% (8/30) did so twice. Similarly, 63% (19/30) of non-cognate learners matched all verbs with their pictures in one attempt, 33% (10/30) did so in two attempts and 3% (1/30) needed a third attempt. Table 4.14 reports cognate and non-cognate learners' mean and median number of attempts at the task, dispersion measures and 95% confidence intervals for the two values. A Wilcoxon rank-sum test indicated that the sample of number of attempts of non-cognate learners did not have significantly larger values than the sample of number of attempts of cognate learners, i.e. there was not a cognate facilitation effect (W = 401, p = .38, r = 0.12).

	Cognate learners	**Non-cognate learners**
Mean	1.27	1.40
SD	0.45	0.56
95%CI	[1.10, 1.43]	[1.20, 1.61]
Median	1	1
MAD	0	0
95%CI	[1,1]	[1,2]

TABLE 4.14. Information regarding the number of attempts at the first picture-verb matching task in Experiment 5 for cognate and non-cognate learners. This includes the mean, Standard Deviation (SD) and the 95% Confidence Interval (95%CI) of the mean, the median, the Median Absolute Deviation (MAD) and the 95%CI of the median.

5.7.1.2. Picture-naming task

The picture-naming task was performed until all pictures of nouns and verbs were named correctly. Cognate and non-cognate learners had to produce the same five non-cognate nouns and four either cognate or non-cognate verbs.

<u>Noun picture naming</u>

Cognate and non-cognate learners needed 1-3 attempts to name all pictures of nouns accurately. The two groups' mean and median number of attempts, dispersion measures and 95% confidence intervals for the two values are presented in Table 4.15. While the mean number of attempts was greater for non-cognate learners than for cognate learners, the two groups had equal medians. Additionally, the distributions of number of attempts of cognate and non-cognate learners had comparable shapes, i.e. one did not have significantly larger values than the other one (W = 365, p = .15, r = 0.19).

	Cognate learners	Non-cognate learners
Mean	1.63	1.87
SD	0.56	0.63
95%CI	[1.43, 1.84]	[1.63, 2.10]
Median	2	2
MAD	0	0
95%CI	[1,2]	[2,2]

TABLE 4.15. Information regarding the number of attempts at the noun picture-naming task in Experiment 5 for cognate and non-cognate learners. This includes the mean, Standard Deviation (SD) and the 95% Confidence Interval (95%CI) of the mean, the median, the Median Absolute Deviation (MAD) and the 95%CI of the median.

Verb picture naming

Cognate learners named all pictures of verbs correctly in 1-2 attempts at the task. Specifically, 87% (26/30) did so in their first attempt and the remaining 13% (4/30), in their second attempt. Conversely, non-cognate learners named all pictures of verbs correctly in 1-3 attempts: 27% (8/30) in their first attempt, 60% (18/30) in their second attempt and 13% (4/30) in their third attempt. The mean and the median number of attempts were larger for non-cognate learners than for cognate learners (Table 4.16). Crucially, the distribution of number of attempts of non-cognate learners had significantly larger values than the distribution of number of attempts of cognate learners, i.e. there was a cognate facilitation effect ($W = 172$, $p < .001$, large effect size of $r = 0.61$).

	Cognate learners	Non-cognate learners
Mean	1.13	1.87
SD	0.35	0.63
95%CI	[1.00, 1.26]	[1.63, 2.10]
Median	1	2
MAD	0	0
95%CI	[1,1]	[2,2]

TABLE 4.16. Information regarding the number of attempts at the first verb picture-naming task in Experiment 5 for cognate and non-cognate learners. This includes the mean, Standard Deviation (SD) and the 95% Confidence Interval (95%CI) of the mean, the median, the Median Absolute Deviation (MAD) and the 95%CI of the median.

5.7.2. Second vocabulary-learning phase

Cognate and non-cognate learners learnt the same four non-cognate verbs. The picture-word matching task and the picture-naming task were performed until 100% accuracy.

5.7.2.1. Picture-word matching task

Cognate and non-cognate learners correctly matched all verbs with their pictures in 1-3 attempts. The two groups attempted the task the same mean and median number of times (Table 4.17). The distributions of number of attempts of cognate and non-cognate learners had comparable shapes, i.e. none of the distributions had significantly larger values than the other one ($W = 439$, $p = .83$, $r = 0.03$).

	Cognate learners	Non-cognate learners
Mean	1.27	1.27
SD	0.58	0.52
95%CI	[1.05, 1.48]	[1.07, 1.46]
Median	1	1
MAD	0	0
95%CI	[1,1]	[1,1]

TABLE 4.17. Information regarding the number of attempts at the second picture-verb matching task in Experiment 5 for cognate and non-cognate learners. This includes the mean, Standard Deviation (SD) and the 95% Confidence Interval (95%CI) of the mean, the median, the Median Absolute Deviation (MAD) and the 95%CI of the median.

5.7.2.2. Picture-naming task

Cognate and non-cognate learners needed either one or two attempts to name all pictures of verbs correctly, except for a non-cognate learner who attempted the task a third time. The two groups' mean and median number of attempts, dispersion measures and 95% confidence intervals for the two values are reported in Table 4.18. The mean number of attempts was larger for non-cognate learners than for cognate learners, but the two groups had equal medians. The distributions of number of attempts of the two groups of learners had comparable values ($W = 417$, $p = .51$, $r = 0.09$).

	Cognate learners	Non-cognate learners
Mean	1.20	1.30
SD	0.41	0.53
95%CI	[1.05, 1.35]	[1,10, 1.50]
Median	1	1
MAD	0	0
95%CI	[1,1]	[1,1]

TABLE 4.18. Information regarding the number of attempts at the second verb picture-naming task in Experiment 5 for cognate and non-cognate learners. This includes the mean, Standard Deviation (SD) and the 95% Confidence Interval (95%CI) of the mean, the median, the Median Absolute Deviation (MAD) and the 95%CI of the median.

5.7.3.Testing phase

5.7.3.1.Sentence-picture congruency task

Performance on semantically incongruent sentence-picture pairs

Cognate learners correctly judged a mean of 87.14% (*SD* = 33.51%, 95%CI = [84.41, 89.86]) of all semantically incongruent pairs with agent violation and a mean of 86.04% (*SD* = 34.69%, 95%CI = [83.19, 88.88]) of all semantically incongruent pairs with patient violation. Similarly, non-cognate learners correctly judged a mean of 85.10% (*SD* = 35.64%, 95%CI = [82.18, 88.01]) of all semantically incongruent pairs with agent violation and a mean of 84.44% (*SD* = 36.27%, 95%CI = [81.50, 87.39]) of all semantically incongruent pairs with patient violation. Accuracy did not significantly vary as a function of Group of learners (β = 0.05, *SE* = 0.23, *z* = 0.23, *p* = .82), Type of sentence-picture pair (β = 0.24, *SE* = 0.15, *z* = 1.61, *p* = .11) or the interaction between the two (β = -0.03, *SE* = 0.26, *z* = -0.12, *p* = .90). That the two groups detected a mismatch between one of the nouns in the sentence and the ones in the picture in over 84% of the pairs on average further demonstrates that learning of nouns was very good. The results also confirm that this lexical learning was comparable for cognate and non-cognate learners.

Performance on syntactically congruent and incongruent sentence-picture pairs

The structures as part of the linguistic system

Table 4.19 and Figure 4.9 summarise cognate and non-cognate learners' mean accuracy percentages when judging syntactically congruent SVO sentence-picture pairs (SVO_{congr}), incongruent SVO sentence-picture pairs ($SVO_{incongr}$), congruent OVS sentence-picture pairs (OVS_{congr}) and incongruent OVS sentence-picture pairs ($OVS_{incongr}$). The two groups judged as required significantly more than 50% of all congruent and incongruent SVO sentence-picture pairs and accuracy was high (above 80%) in all cases. The two groups judged significantly more than 50% of all congruent OVS sentence-picture pairs correctly, but they could not reject significantly more than 50% of all incongruent OVS sentence-picture pairs (cognate learners, *t* (29) = 1.28, *p* = .11, *d* = 0.23; non-cognate learners, *t* (29) = 0.92, *p* = .18, *d* = 0.17).

		SVO_{congr}	**$SVO_{incongr}$**	**OVS_{congr}**	**$OVS_{incongr}$**
Cognate learners	*M*	94.18***	81.34***	85.66***	58.16
	SD	23.46	39.03	35.11	49.42
	95%CI	[91.48, 96.88]	[76.78, 85.90]	[81.58, 89.75]	[52.36, 63.95]
Non-cognate learners	*M*	93.79***	86.62***	65.28*	56.99
	SD	24.17	34.10	47.69	49.60
	95%CI	[91.00, 96.59]	[82.64, 90.60]	[59.75, 70.81]	[51.14, 62.83]

TABLE 4.19. Cognate and non-cognate learners' mean accuracy (%), standard deviations (%) and 95% confidence intervals in syntactically congruent and incongruent SVO and OVS sentence-picture pairs in Experiment 5. *M* = Mean, *SD* = Standard Deviation, *95%CI* = 95% Confidence Interval. Significance from chance: * *p* < .05. *** *p* < .001.

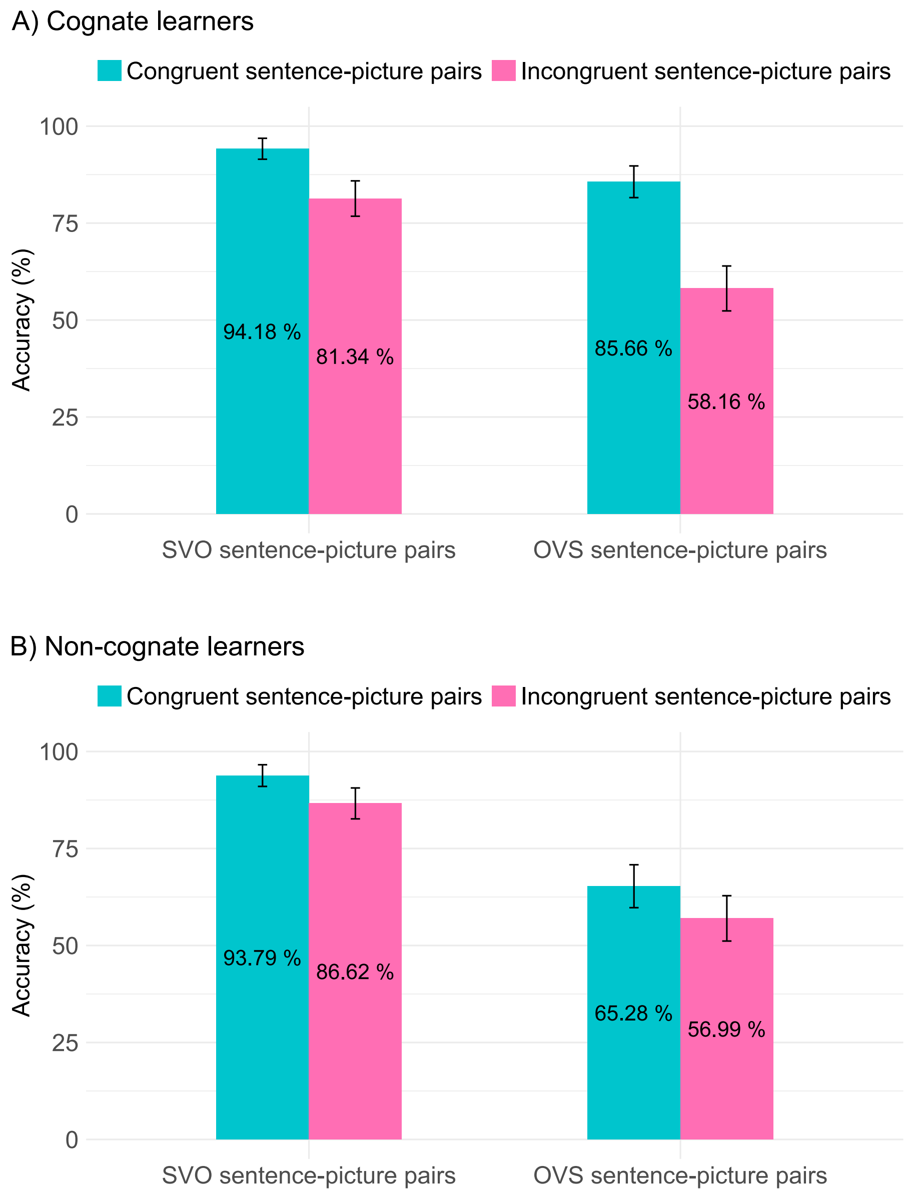


FIGURE 4.9. Mean accuracy (%) in syntactically congruent and incongruent SVO and OVS sentence-picture pairs for A) cognate learners and B) non-cognate learners in Experiment 5. Error bars represent 95% confidence intervals.

On the one hand, these results suggest that the SVO structure with prepositional patient marking was part of cognate and non-cognate learners' linguistic system. In this light, I collapsed accuracy for congruent and incongruent SVO sentence-picture pairs and assessed whether overall knowledge of the structure was comparable for cognate and non-cognate learners. Cognate learners correctly judged a mean of 87.85% (*SD* = 32.70%, 95%CI = [85.17, 90.52]) of all SVO sentence-picture pairs. Mean accuracy was slightly higher for non-cognate learners, 90.24% (*SD* = 29.70%, 95%CI = [87.81, 92.68]). A generalized linear mixed effects model indicated that the difference between groups was not significant (β = 0.31, *SE* = 0.36, *z* = 0.85; *p* = .40). The estimated logit coefficient of the effect of group of learners on accuracy was 0.31, which corresponds to an odds ratio of 1.36 to 1. That is, the odds of judging an SVO sentence-picture pair correctly as opposed to incorrectly were 1.36 times higher for non-cognate learners than for cognate learners, but this effect was not significant. This suggests that the SVO structure with prepositional patient marking was comparably established in cognate and non-cognate learners' linguistic system, in line with Hypothesis 1a (H1a) and Hypothesis 2a (H2a) postulated for this experiment.

On the other hand, there is no evidence that the OVS structure with prepositional patient marking was established neither in cognate learners' nor in non-cognate learners' linguistic system. The mean index of response bias *c* indicated that both groups tended to judge as congruent those sentence-picture pairs in which *a* preceded the first noun in the sentence, irrespective of whether this noun was the agent or the patient in the picture. Specifically, this mean index was negative and significantly below zero both for cognate learners (M = -0.38, SD = 0.40; t (29) = -5.29, p <.001, d = -0.97) and for non-cognate learners (M = -0.15, SD = 0.35; t (29) = -2.29, p = .01, d = -0.42). Descriptively, the two groups judged a similar percentage of syntactically incongruent OVS sentence-picture pairs correctly. By contrast, accuracy when judging syntactically congruent OVS sentence-picture pairs was much higher for cognate learners than for non-cognate learners. This was confirmed statistically. A generalized linear mixed effects model yielded an interaction between the effect of Group of learners and Congruency of the sentence-picture pair on accuracy (β = 1.37, SE = 0.36, z = 3.85; p < .001). Post-hoc comparisons indicated that the two groups' accuracy significantly differed when judging syntactically congruent OVS sentence-picture pairs (β = 1.55, SE = 0.62, z = 2.51; p = .01), but not when judging syntactically incongruent ones (β = -0.35, SE = 0.74, z = -0.47; p = .64). The fact that cognate learners correctly judged significantly more congruent OVS sentence-picture pairs than non-cognate learners might suggest that the former found it easier to identify that the sentence-initial *a* was a patient mark. However, this might also reflect that cognate learners had a stronger response bias than non-cognate learners. An explanation for this bias may be found in the strategies that participants used to perform the task, as will be shown below. This information was elicited in the debriefing phase. Nevertheless, it is reported in this section to facilitate the interpretation of participants' strategies together with the results of the sentence-picture congruency task.

<u>Strategies used to perform the sentence-picture congruency task</u>

Participants reported using several strategies in the congruency task. Some claimed that they applied the patient marking knowledge learnt during the exposure phase (70%, 21/30 of all cognate learners and 57%, 17/30 of all non-cognate learners). Others reported relying on their intuition (17%, 5/30 of all cognate learners and 37%, 11/30 of all non-cognate learners). Importantly, 13% (4/30) of all cognate learners and 7% (2/30) of all non-cognate learners said that they judged the congruency of sentence-picture pairs based on whether the nouns in the sentence and the ones in the picture coincided. That is, if the nouns in the sentence and the picture matched —as in syntactically congruent and incongruent sentence-picture pairs— that pair was judged as "correct". Otherwise, the pair was judged as "incorrect". Three out of these four cognate learners reported that *a* was a patient mark in the debriefing phase and reported noticing this during the exposure phase (see Section 5.7.4). This suggests that these learners did not use their syntax knowledge in the sentence-picture congruency task despite having learnt the target structures. In addition, it could be that the rest of learners who consciously followed this strategy did not report the function of *a,* but knew it unconsciously.

In sum, the strategy to judge sentence-picture pairs based on vocabulary only, which some cognate and non-cognate learners followed, affected these learners' accuracy when judging OVS sentence-picture pairs and probably also when judging SVO sentence-picture pairs. In spite of this, overall cognate and non-cognate learners clearly differentiated syntactically congruent SVO sentence-picture pairs from their syntactically incongruent counterparts. On the one hand, the OVS structure could be established in learners' linguistic system, but this could be masked by the bias towards judging sentence-picture pairs as "correct". On the other hand, the fact that learning of the SVO structure was not masked by this response bias suggests that learning of the OVS structure, if any, may have been weaker. In other words, the SVO structure may have been more robustly established in cognate and non-cognate learners' linguistic system than the OVS structure, as claimed by Hypothesis 1b (H1b) and Hypothesis 2b (H2b). Yet, the response bias decreases the reliability of the results of the sentence-picture congruency task.

5.7.3.2. Production task

The structures as part of the linguistic system

Cognate learners wrote 70.42% (169/240) of SVO picture descriptions (i.e. with the agent or subject in the picture in the first position of the sentence) and 29.58% (71/240) of OVS picture descriptions (i.e. with the patient or object in the picture in the first position of the sentence). Similarly, non-cognate learners wrote 79.58% (191/240) of SVO picture descriptions and 20.42% (49/240) of OVS picture descriptions. The fact that both groups of learners produced the two word orders indicates that they were aware that the agent/subject and the patient/object could appear in sentence-initial position. Both groups produced more SVO sentences than OVS ones and, descriptively, the proportion of subject-initial and object-initial sentences was similar in the two groups. Table 4.20 provides a descriptive analysis of cognate and non-cognate learners' verb-medial picture descriptions. First, it indicates the percentage of sentences in which the two groups used *a* before a noun (*a*-marking). Cognate and non-cognate learners used *a*-marking in significantly more than 50% of their productions. This suggests that, overall, both groups were aware that the sentences of the mini-language contained the word *a.* Cognate learners produced a higher percentage of picture descriptions with *a*-marking than non-cognate learners did. However, the difference between groups was not significant (β =-0.68, SE = 5.23, z = -0.13, p = .90). Second, Table 4.20 shows the percentage of picture descriptions with *a*-marking in which *a* was written before the first noun in the sentence or before the second noun in the sentence. Cognate and non-cognate learners had an almost identical preference for writing sentences with *a*-marking before the second noun. Yet, they also used *a*-marking before the first noun, which indicates that both groups knew that *a* could appear in these two positions. Finally, in the rightmost column of the table I report the percentage of sentences in which *a* was used correctly before the first noun and before the second noun (i.e. verb-medial sentences with correct patient marking). Cognate and non-cognate learners used patient marking correctly in significantly more than 50% of picture descriptions with *a* before the second noun, which suggests that the SVO structure with prepositional patient marking was part of learners' linguistic system. Likewise, both groups used patient marking correctly in significantly more than 50% of picture descriptions with *a* before the first noun, indicative of the fact that the OVS structure with prepositional patient marking was also established in learners' linguistic system.

Cognate learners	*Picture descriptions with a-marking*	95.00%*** (228/240)	*a-marking before the first noun*	28.51% (65/228)	*Correct patient marking*	89.23*** (58/65)
Cognate learners	*Picture descriptions with a-marking*	95.00%*** (228/240)	*a-marking before the second noun*	71.49% (163/228)	*Correct patient marking*	92.02%*** (150/163)
Non-cognate learners	*Picture descriptions with a-marking*	90.42%*** (217/240)	*a-marking before the first noun*	28.23% (59/209)[23]	*Correct patient marking*	76.27%*** (45/59)
Non-cognate learners	*Picture descriptions with a-marking*	90.42%*** (217/240)	*a-marking before the second noun*	71.77% (150/209)	*Correct patient marking*	98.00%*** (147/150)

TABLE 4.20. Cognate and non-cognate learners' percentages of (i) verb-medial picture descriptions with *a*-marking, (ii) with *a*-marking before the first noun and before the second noun and (iii) with correct patient marking in Experiment 5. Significance from chance: *** $p < .001$.

Comparing the establishment of the structures in cognate and non-cognate learners' linguistic system

A generalized linear mixed effects model tested whether cognate and non-cognate learners differed in the accuracy with which they wrote verb-medial sentences with *a* before the first noun and before the second noun. That is, the model assessed whether there was an interaction between Group of learners (Cognate vs. Non-cognate) and Position of *a* (Before the first noun vs. Before the second noun) on sentence accuracy. The test yielded no effect of Group of learners (β = -0.64, SE = 2.18, z = -0.30, p = .77), an effect of Position of *a* (β = -2.35, SE = 1.17, z = -2.02, p = .04) and no interaction between the two variables (β = 3.31, SE = 2.21, z = 1.50, p = .13). The lack of a significant group effect suggests that cognate and non-cognate learners were comparably accurate when writing verb-medial sentences with *a*. This is taken as evidence that, overall, the target structures were comparably established in cognate and

[23] The eight picture descriptions of a non-cognate learner were removed because s/he wrote *a* before the two nouns in the sentence.

non-cognate learners' linguistic system, as claimed by H1a. Learners wrote a mean of 95.19% (*SD* = 37.54%, 95%CI = [92.81, 97.58]) of correct picture descriptions with *a*-marking before the second noun (collapsing picture descriptions of cognate and non-cognate learners). The mean percentage of correct picture descriptions with *a*-marking before the first noun was lower, 83.20% (*SD* = 21.43%, 95%CI = [76.55, 89.85]). The effect of Position of *a* suggests that, overall, learners were significantly more accurate in their verb-medial descriptions with *a* before the second noun than before the first noun. The absence of a significant interaction indicates that this effect did not depend on the group of learners. This suggests that the SVO structure was more robustly established than the OVS structure in learners' linguistic system and that this difference was comparable for cognate and non-cognate learners, as claimed by H1b.

5.7.4. Debriefing phase

In this section, I summarise participants' responses to the questions in the verbal report assessing awareness of patient marking. Responses to the question about the strategies used in the sentence-picture congruency task were reported together with the results of that task (see Section 5.7.3). All cognate learners reported that the sentences of the experiment contained an *a*. Of these learners, 47% (14/30) reported that *a* was placed before the person who received the action of the verb (the patient of the sentence). Additionally, 27% (8/30) did not verbalize the function of *a,* but correctly exemplified how it was used in SVO and OVS sentences using the equivalent constructions from Spanish (e.g. *La a aparecía en dos tipos de frases, una del tipo: El médico saluda* ***a****l actor y la otra del tipo:* ***A****l actor lo saluda el médico; The a appeared in two types of sentences, one of the type: The doctor greets* ***a*** *the actor [SV-a-O order] and the other of the type:* ***A*** *the actor greets the doctor [a-OVS order],* P4)[24]. On the other hand, 10% (3/30) of cognate learners did not explicitly refer to *a* as a patient mark nor used example sentences, but provided answers which evidenced that they knew how *a* was used. Specifically, P2 reported that *a* could appear in two different positions in the sentence and claimed that the function of this word was the same as in Spanish, which I interpret as referring to it being a patient mark. P5 reported that when *a* appeared at the beginning of the sentence, this was in the passive voice. I interpret this as referring to the fact that the patient was in first position and, thus, followed *a*, and that the agent appeared later in the sentence. P23 reported that when *a* appeared between the first noun and the second noun in the sentence, the meaning expressed was that the first noun did something to the second. By contrast, when *a* appeared at the beginning of the sentence, the doer of the action was the second noun. The remaining 17% (5/30) of cognate learners did not verbalize the function of *a* (P1 and P22), incorrectly reported that it marked the agent of the sentence (P14) or reported that it was either an agent or a patient mark, but did not specify which (P19 and P24).

Turning to non-cognate learners, all but one reported that the experimental sentences included *a.* Out of these learners, 48% (14/29) stated that *a* appeared before the person who

[24] P = Participant. SVO and OVS sentences meaning "The doctor greets the actor".

received the action of the verb or who did not perform the action (the patient). Additionally, 7% (2/29) of learners correctly specified the function of *a,* but only in SVO sentences. On the other hand, 14% (4/29) of learners did not verbalize how *a* was used*,* but correctly exemplified this using SVO and OVS sentences in Spanish. Finally, a learner (P20) exemplified how *a* was used in SVO sentences and mentioned that when *a* appeared in sentence-initial position, this was in the passive voice. Like for cognate learners, I interpret this as knowledge that the patient was in the first position of the sentence and followed *a*. As for the remaining 28% (8/29) of non-cognate learners, they did not state the function of *a* (P14 and P17), reported that it was an article (P7), that it marked a connection between the verb and a noun (P8) or that it could be an agent mark or a patient mark (P9, P16, P25 and P28).

To sum up, 83% (25/30) of all cognate learners and 70% (21/30) of all non-cognate learners provided answers in the verbal report that revealed that they knew the function of *a.* I considered these learners *aware* of patient marking. I considered the remaining 17% (5/30) of cognate learners and 30% (9/30) of non-cognate learners *unaware* of this marking, since they either did not report that sentences contained the word *a* or did not state how *a* was (correctly) used. Aware learners wrote more than 50% of correct picture descriptions in the production task, evidencing syntax learning. They all reported noticing how *a* was used during the exposure phase. Conversely, unaware learners did not write more than 50% of correct picture descriptions.

5.7.5. Reading span task

Cognate learners had a mean partial reading span score of 44.45 (*SD* = 11.76, 95%CI = [39.98, 48.92]). Non-cognate learners had a mean score of 48.90 (*SD* = 10.47, 95%CI = [44.99, 52.81]). The two groups' scores were statistically comparable (t (57) = -1.54, p = .13, d = -0.40). This suggests that cognate and non-cognate learners had similar working memory capacities.

5.8. Discussion

Experiment 5 examined whether the facilitative role of cognates in the acquisition of L1-L2 dissimilar structures found in Experiment 4 varied for L1-L2 similar structures. Specifically, the experiment explored how Spanish natives with no knowledge of Basque learnt two syntactic constructions based on Spanish grammar —SVO and OVS structures with prepositional patient marking— by being exposed to these structures in sentences which contained two Spanish-Basque non-cognate nouns and either a cognate verb (cognate learners) or a non-cognate verb (non-cognate learners). Like in Experiment 4, first, cognate and non-cognate learners learnt the vocabulary via word-picture pairs. Then, they were exposed to sentences formed by the L2 structures and containing the vocabulary learnt, each paired with a picture. After that, they learnt novel non-cognate verbs and were tested on the structures with these verbs in a sentence-picture congruency task and in a picture-description task. The last part of the experiment was a verbal report. I postulated two possible hypotheses about the influence that cognates could have on the acquisition of the cross-linguistically similar L2 structures, each having two parts. These were based on how I proposed that cognate and non-cognate words could be stored and processed within MOGUL, how this framework could account for acquisition by processing of L1-L2 similar structures and how it conceives the interaction between lexical and syntactic processing. The picture-description task provided evidence in favour of the first of my hypotheses, subdivided into Hypothesis 1a (H1a) and Hypothesis 1b (H1b).

H1a claimed that, overall, cognates would not facilitate the acquisition of the L2 structures. On the one hand, I proposed that cognates would be more activated than non-cognates during processing and, thus, that a stronger activation would spread from the cognate verb than from the non-cognate verb to the structures containing them. Consequently, the structures processed with cognates would be more strongly activated than the ones processed with non-cognates. On the other hand, I proposed that the L2 structures would be processed by accessing similar L1 constructions firmly stored in learners' linguistic system by the time that L2 acquisition started. This would cause that, even if the structures were slightly more activated when processed with cognates than with non-cognates, they ended up being comparably established in cognate and non-cognate learners' linguistic system. In support of H1a, cognate and non-cognate learners were comparably accurate when writing sentences with verb-medial word order and prepositional patient marking in the picture-description task. Turning to H1b, it claimed that there would be a learning advantage for the SVO structure over the OVS one and that this advantage would be similar for cognate and non-cognate learners. This is because the L1 SVO structure is more frequent than the L1 OVS structure (cf. ADESSE corpus, García-Miguel et al., 2010) and, thus, the former would be more firmly established in learners' linguistic system than the latter. In support of H1b, learners were significantly more accurate when writing verb-medial picture descriptions with prepositional patient marking before the second noun in the sentence (SVO structure) than before the first noun in the sentence (OVS structure). This effect did not vary as a function of group of learners. In the next sections, I discuss the results of the first vocabulary-learning phase (focusing on cognate and non-cognate verbs), the testing phase and the debriefing phase.

5.8.1. Discussion of the first vocabulary-learning phase

Vocabulary learning was assessed in a picture-word matching task and in a picture-naming task. The two tasks were conducted until 100% accuracy, so that cognate and non-cognate learners mastered all non-cognate nouns and all cognate or non-cognate verbs to the same degree. In the picture-word matching task, cognate and non-cognate learners matched all nouns with their pictures in a statistically similar number of attempts. The same occurred when matching cognate vs. non-cognate verbs with their pictures. By contrast, in the picture-naming task, the two groups named all nouns correctly in a comparable number of attempts, but cognate learners named all verbs correctly in significantly fewer attempts than non-cognate learners did. This suggests that cognate verbs were learnt faster than non-cognate verbs. This cognate facilitation effect adds to the large number of studies supporting this effect, including Experiment 4 in this chapter (e.g. Antón & Duñabeitia, 2020; Comesaña et al., 2019; Comesaña, Soares, et al., 2012; de Groot & Keijzer, 2000; Lotto & de Groot, 1998; Marecka et al., 2021; Tonzar et al., 2009; Valente et al., 2018; see Section 3.8.1 for an overview).

In Experiment 4, I discussed how various models of vocabulary learning could account for this facilitation effect, including the Parasitic Model of vocabulary development (Ecke & Hall, 1998; Hall, 1996, 2002, and more), the RHM (Kroll & Stewart, 1994) and the BIA-d model (Grainger et al., 2010) (see Section 3.8.1 for a full account). I also discussed how I proposed that cognate and non-cognate verbs were processed and learnt during the vocabulary-learning phase within MOGUL. In short, I assumed that because cross-linguistically similar words are co-activated during word processing, cognate verbs were more strongly activated than non-cognate verbs each time that they were processed. This caused that, after cognate and non-cognate learners were exposed to cognate and non-cognate verbs (prior to conducting the picture-word matching task and the picture-naming task), cognates had a higher resting activation level than non-cognates and, thus, were more firmly established in the linguistic system. This reasoning could also hold in Experiment 5 and would explain why, in the picture-naming task, cognate learners named all pictures of verbs appropriately in significantly fewer attempts than non-cognate learners did. However, this reasoning would not explain why, in the picture-word matching task, cognate and non-cognate learners correctly matched all cognate and non-cognate verbs with pictures in a similar number of attempts. I argue that there may be at least two explanations for this null effect.

First, although in the picture-word matching task learners were asked to respond as quickly as possible, there was actually no time limit. This allowed both groups of learners to think about their responses carefully before matching each verb with a picture and could have helped non-cognate learners make less mistakes than they probably would have made if there had been time pressure. In addition, in that case reaction times could have determined whether cognate verbs were matched with pictures significantly faster than non-cognate verbs, in line with previous studies finding a cognate facilitation effect in reaction times but not in accuracy (e.g. Valente et al., 2018). Second, non-cognate learners could have followed one or more strategies to remember which verb was paired with each picture, e.g. remembering just the first letters of the verb, instead of the whole word. This could have helped them achieve 100% accuracy in the picture-word matching task and would explain why non-cognate learners had more difficulty than cognate learners did when they had to produce the verbs in the picture-naming task, an exercise that is more difficult than matching words with pictures.

5.8.2. Discussion of the testing phase

5.8.2.1. Sentence-picture congruency task

Cognate and non-cognate learners judged SVO and OVS sentence-picture pairs as congruent or incongruent. Half of these pairs were semantically incongruent, i.e. one of the nouns in the sentence did not coincide with the ones in the picture. The other half were syntactically congruent (the noun marked as the patient in the sentence was the patient in the picture) or incongruent (the noun marked as the patient in the sentence was the agent in the picture). Both groups judged most semantically incongruent sentence-picture pairs correctly. Likewise, accuracy when judging syntactically congruent and incongruent SVO sentence-picture pairs was above chance and, overall, accuracy was not significantly higher for cognate learners than for non-cognate learners. This suggests that the SVO structure with prepositional patient marking was comparably established in cognate and non-cognate learners' linguistic system. By contrast, both groups' accuracy when judging syntactically congruent OVS sentence-picture pairs was above chance, but it was not when judging syntactically incongruent OVS pairs. In other words, there was a bias towards judging OVS sentence-picture pairs as congruent. This was partly the consequence of some learners consciously judging the congruency of sentence-picture pairs only by comparing the nouns in the sentence and the picture (irrespective of patient marking). Learners following this strategy should have used it to judge SVO sentence-picture pairs too, so this response bias probably does not entirely account for the difference in accuracy between SVO and OVS sentence-picture pairs. Yet, it diminishes the reliability of the congruency task. In what follows, I discuss how cognate and non-cognate learners could have processed SVO and OVS sentences during the exposure phase and how this could have yielded the results observed.

I will first review how I hypothesized that Spanish natives without knowledge of Basque would process SVO sentences with prepositional patient marking and with a Spanish-Basque cognate or non-cognate verb (cf. Section 4.2). I assumed that when learners encountered an SVO sentence for the first time, the syntactic processor would process it using a syntactic representation such as $[_{CP}$ NP $[_{VP}$ V $[_{PP}$ P NP]]], stored in learners' linguistic system during L1 acquisition. I argued that due to the agent-first preference and the fact that in Spanish the preposition *a* is placed before animate and specific direct objects/patients, the first NP in the sentence would be associated with the conceptual role of AGENT and the second NP, with the conceptual role of PATIENT. This would lead to a target-like processing of the sentence and to an increase in its structure's resting activation level. Importantly, I hypothesized that processing SVO sentences with a cognate verb vs. a non-cognate verb would not significantly affect the establishment of the SVO structure in learners' linguistic system. Specifically, I proposed that since cognate verbs are more strongly activated than non-cognate verbs, the SVO structure processed with a cognate would be more activated than the same structure processed with a non-cognate. Consequently, the increase in the structure's resting activation level would be higher when a sentence included a cognate verb than a non-cognate verb. Yet, since the structure was repeatedly processed as part of the L1, it would already have a very high resting activation level and the increases in this resting level derived from processing would be small. Thus, by the end of the exposure phase any difference in resting activation level resulting from processing the structure with a cognate verb as opposed to a non-cognate verb would be non-significant. This outcome of the exposure phase would explain cognate and non-cognate learners' performance in the sentence-picture congruency task. On the one hand, since the SVO structure was present in learners' linguistic system, both groups judged most syntactically congruent and incongruent SVO sentence-picture pairs correctly. On the other hand, since the SVO structure was comparably established in cognate and non-cognate learners' system, overall accuracy in SVO sentence-picture pairs did not statistically differ between groups.

I now review how I hypothesized that Spanish natives without knowledge of Basque would process OVS sentences with prepositional patient marking (cf. Section 4.2). Since in the congruency task there was no evidence that the OVS structure was established neither in cognate nor in non-cognate learners' linguistic system, I will not discuss how I hypothesized that OVS sentences would be processed with a cognate verb vs. a non-cognate verb. I argued that when learners read and listened to an OVS sentence for the first time, the syntactic processor would process it using a syntactic representation such as [$_{CP}$ [$_{NP}$ D N] [$_{VP}$ V [$_{PP}$ P NP]]], stored in learners' linguistic system as part of the L1[25]. This is because in Spanish it is more likely to encounter an NP than a PP in preverbal position. I argued that following the agent-first preference, the first NP in the sentence would be interpreted as the AGENT and the second NP, as the PATIENT. This would lead to an incorrect interpretation of the sentence, as indicated by the picture accompanying it. This misprocessing of OVS sentences would eventually cause learners to reanalyse them in terms of an object/patient-first structure with prepositional patient marking. In subsequent processing occasions, an L1 structure such as [$_{CP}$ [$_{PP}$ P NP] [$_{VP}$ (CL) V NP]] with the first NP associated with the conceptual role of PATIENT would be used to process OVS sentences and its resting activation level would increase.

I claim that some cognate and non-cognate learners processed OVS sentences during the exposure phase as described. This caused that, in the sentence-picture congruency task, they judged syntactically congruent and incongruent OVS sentence-picture pairs as required. Yet, there was an overall bias to judge both types of OVS sentence-picture pairs as congruent. In what follows, I propose two explanations for this bias. First, it could be that some learners processed OVS sentences as hypothesized, but that they found the congruency task too demanding. As mentioned for Experiment 4 (Section 3.8.2.1), after being exposed to a sentence-picture pair for about a second, learners had 5 seconds to decide whether the sentence and the picture were congruent or not. In this time, they had to process the sentence (accessing its lexical items and integrating them into a syntactic structure), interpret it and compare this interpretation against the picture. Even if in Experiment 5 sentences could be processed using compatible L1 structures, performing congruency judgements in the time allowed might have been quite costly, particularly for OVS sentence-picture pairs, which were processed accessing a structure that is not very frequent in the L1. Consequently, even if the target structures were learnt, some learners could have decided to make congruency judgements just based on whether or not the nouns in the sentence matched the ones in the picture. This strategy, which some learners verbalized in the debriefing phase, would have caused that they judged as congruent both syntactically congruent and incongruent sentence-

[25] The syntactic representation of the preposition could be coindexed with an orthographically and phonologically null representation.

picture pairs. Second, I propose that some learners may have processed OVS sentences as hypothesized, i.e. as SVO, first, and then reanalyse them in terms of an object-first structure. Yet, this reanalysis may have occurred towards the end of the exposure phase. Hence, the amount of sentences processed using the OVS structure may not have been enough for it to be activated and consistently used during the congruency task. Consequently, learners could have simply judged most OVS sentence-picture pairs as congruent.

5.8.2.2. Production task

Cognate and non-cognate learners saw pictures of transitive actions and were asked to write a sentence describing each picture using the nouns and the verbs learnt. As indicated, the results of this task revealed that overall learning of the cross-linguistically similar L2 structures, with SVO or OVS word order and prepositional patient marking, was comparable for cognate and non-cognate learners. Both groups were comparably accurate when producing verb-medial sentences with the preposition *a* before one of the nouns in the sentence. In the previous section, I discussed how I proposed that cognate and non-cognate learners processed SVO sentences with prepositional patient marking from the moment they first encountered them in the exposure phase. In short, I assumed that these sentences were processed using an L1 structure having a high resting activation level, which increased even more each time that an SVO sentence was processed. I proposed that the stronger activation of a cognate verb, compared to a non-cognate verb, spread to the SVO structure processed, causing it to be more activated when including a cognate than a non-cognate. As a result, I argued that the increase in the structure's resting activation level derived from processing SVO sentences was higher when these contained a cognate verb than a non-cognate verb. However, due to the structure's initially high resting activation level, I claimed that by the end of the exposure phase this level was comparable for cognate and non-cognate learners (i.e. the structure was comparably established in the two groups' linguistic system).

I also discussed how I proposed that, overall, learners processed OVS sentences with prepositional patient marking from first exposure. In brief, I assumed that these sentences were processed using an L1 syntactic representation in which the first and the second NPs were incorrectly coindexed with the conceptual roles of AGENT and PATIENT, respectively. After misprocessing a few OVS sentences, these were reanalysed in terms of an object/patient-first structure using a compatible L1 syntactic representation. The resting activation level of this structure increased each time that an OVS sentence was processed. I now turn to the role of cognates in processing OVS sentences (cf. Section 4.2). The results of the production task suggest that the effect of cognates was similar to the one described for SVO sentences. That is, the L1 OVS structure used to process L2 OVS sentences was stored in the linguistic system with a quite high resting activation level (although not as high as that of the L1 SVO structure, see below). Since the OVS structure was firmly established in learners' linguistic system, overall increases in its resting activation level were small. Consequently, the resting activation level of the structure processed with cognates was non-significantly higher than that of the structure processed with non-cognates (i.e. the structure was comparably established in cognate and non-cognate learners' linguistic system).

In sum, the fact that learning of the target structures was similar for cognate and non-cognate learners would explain why the two groups were equally accurate when writing verb-medial picture descriptions with patient marking. This result aligns with some of the results of Chapter 3. Specifically, in Chapter 3 I showed that the stronger activation of high frequency verbs compared to low frequency verbs did not facilitate the acquisition of an L1-L2 similar structure. Similarly, in Experiment 5 I showed that the stronger activation of cognate verbs compared to non-cognate verbs did not facilitate the acquisition of L1-L2 similar structures.

In addition, in the production task learners were significantly more accurate when writing verb-medial picture descriptions with *a* before the second noun in the sentence (i.e. correct SVO sentences with prepositional patient marking) than before the first noun in the sentence (i.e. correct OVS sentences with prepositional patient marking). The difference between the two types of sentences did not significantly vary as a function of group of learners. This provides evidence in favour of the fact that the SVO structure was more firmly established in cognate and non-cognate learners' linguistic system than the OVS structure. I hypothesize that this is because the frequency of occurrence of the L1 SVO structure used to process L2 SVO sentences was higher than that of the L1 OVS structure used to process L2 OVS sentences (e.g. ADESSE corpus, see Section 4.2). Hence, the resting activation level of the first structure was also higher than that of the second.

In the previous section, I also discussed the possibility that for some learners the OVS structure with patient marking started to be used for processing towards the end of the exposure phase. This could have made it more difficult for the OVS structure to be activated and used during the production task and would explain why some learners wrote incorrect picture descriptions with *a* before the first noun in the sentence, i.e. sentences in which the noun acting as agent/subject in the picture was in sentence-initial position preceded by *a.* Finally, both cognate and non-cognate learners wrote a larger number of sentences with an SVO order (i.e. with the agent/subject in the picture in sentence-initial position) than with an OVS order (i.e. with the patient/object in the picture in sentence-initial position), irrespective of patient marking. The proportion of the two types of sentences was similar in the two groups of learners. Regardless of whether the target structures were stored in learners' linguistic system with a higher or with a lower resting activation level, I assumed that all learners realized that the subject/agent and the object/patient could be the first word of the sentence. This would explain why cognate and non-cognate learners wrote picture descriptions in which the subject or the object in the picture was in sentence-initial position. The fact that both groups wrote more subject-initial sentences than object-initial ones could be attributed to the agent-first preference, i.e. the tendency in most languages (including Spanish, learners' L1) to put agents before patients (Dryer, 2013).

5.8.3. Discussion of the debriefing phase

Experiment 5 used an implicit learning paradigm, which normally produces implicit (non-verbalizable) knowledge of the structures learnt (e.g. Kim & Fenn, 2020; Rebuschat, 2009). Cognate and non-cognate learners were not instructed to look for patterns in the sentence-picture pairs presented during the exposure phase. However, they could have done so anyway, particularly if we consider that L2 learners have metalinguistic awareness. In this light, I expected the learning paradigm to result in explicit (verbalizable) knowledge of patient marking, at least for some cognate and non-cognate learners. Accordingly, more than 80% of cognate learners and 70% of non-cognate learners were aware that *a* was a patient mark and reported this in the verbal report. The remaining learners did not report how *a* was (correctly) used and were considered unaware of patient marking. All aware learners wrote more than 50% of correct picture descriptions (with verb-medial word order and prepositional patient marking) in the production task, thus showing syntax learning. These learners reported noticing how patient marking was realized during the exposure phase, which suggests that learning took place during this part of the experiment and not during the testing phase. Unaware learners did not write more than 50% of picture descriptions correctly. This could indicate that the target structures were not learnt. Alternatively, it could be that unaware learners did learn the structures (even if not evidenced in the production task), but that knowledge of these structures was unconscious or that learners lacked the confidence or the ability to put their knowledge into words. As discussed for Experiment 4 (Section 3.8.3), the sentence-picture congruency task could have helped differentiate between these two options. However, since the results of this task were unreliable, it is not possible to establish whether unaware learners had unconscious knowledge of patient marking or whether the experiment did not elicit learning of the structures for these learners.

Before concluding, some responses to the question of how *a* was used in the L2 must be discussed. First, a cognate learner reported that *a* had the same function as in Spanish. The fact that s/he made explicit reference to the native language indicates that s/he was aware of the cross-linguistic similarity between the structures in the mini-language and in Spanish and suggests that, as hypothesized, the structures of the native language may have been highly active during the experiment. Second, a cognate and a non-cognate learner verbalized that when *a* was the first word of the sentence, that sentence was in the passive voice (i.e. with the patient in initial position). The fact that these learners expressed that the patient was put before the agent as in the passive structure may be attributed to the fact that, in Spanish, the passive is much more frequent than the OVS structure. For example, in the ADESSE corpus there are 100 passive transitive sentences and just 47 OVS sentences with *a* preceding the first argument. Hence, the passive might have been easier to recall and report during the debriefing phase. Third, two non-cognate learners verbalized how *a* was used in the SVO structure, but not in the OVS structure. This suggests that, in line with the results of the testing phase, the SVO structure was more robustly established in these learners' linguistic system than the OVS structure, to the point that learners became aware of the former, but not of the latter. Finally, a non-cognate learner reported that the function of *a* was that of an article. This learner's picture descriptions in the production task were sentences in which the first noun in the sentence, preceded by *a,* was the agent or subject in the picture. An explanation for this could be that this learner processed OVS sentences as subject-initial and did not get to reanalyse them in terms of an object-initial structure.

6. Concluding remarks

The experiments in this chapter are the first to explore, to my knowledge, how cognates affect the acquisition of cross-linguistically dissimilar (Experiment 4) and cross-linguistically similar (Experiment 5) L2 structures by complete beginner adult learners. The findings of this chapter contribute to our understanding of how lexical processing, and particularly lexical activation, interacts with syntactic processing during initial L2 syntax acquisition, extending and enriching the insights into this topic presented in Chapter 3. More precisely, Experiments 4 and 5 suggest that the stronger activation of cognate verbs compared to non-cognate verbs eases the acquisition of cross-linguistically dissimilar L2 structures, which have to be learnt from input. Nevertheless, this facilitation is not found for cross-linguistically similar L2 structures, which can be processed using similar L1 structures and for which no new syntactic representations have to be created. These findings resonate with those of Chapter 3, particularly with the fact that the stronger activation of high frequency verbs compared to low frequency verbs facilitated the acquisition of an L1-L2 dissimilar structure, but not of an L1-L2 similar structure. As far as I know, no model or theory of second language acquisition explicitly details how differences in lexical processing as studied in this chapter would affect L2 syntax acquisition, and none can comprehensively account for the findings of Experiments 4 and 5. I argue that my results may be explained within the MOGUL framework, which I propose can account for the cognate facilitation effect in word processing and for how processing cognates and non-cognates may interact with processing and/or acquisition of cross-linguistically similar and dissimilar L2 structures in real time. In sum, in line with Chapter 3, Chapter 4 presents promising evidence in favour of a facilitative role of lexical activation in initial L2 syntax acquisition, which could usefully continue to be explored in future research.

Chapter 5

General conclusions and directions for future research

The major contributions of this thesis, where I investigated the facilitation exerted by cross-linguistic syntactic similarity and lexical processing in adults' initial acquisition of L1-L2 similar and dissimilar structures are the following:

1. I showed that complete beginner learners demonstrate a learning advantage for an L2 structure that also exists in the L1 over an L2 structure that only exists in the L2 (Chapters 2 and 3). This constitutes novel evidence in favour of the facilitative role of cross-linguistic syntactic similarity at the earliest stage of L2 development and it is a strong validation of models and theories of L2 acquisition predicting this facilitation.

2. I showed for the first time greater learning of L1-L2 dissimilar structures when including words that are activated more strongly during processing, namely high frequency words, compared to low frequency words (Chapter 3) and cognates, compared to non-cognates (Chapter 4). By contrast, I demonstrated learning of L1-L2 similar structures to be comparable with high frequency words vs. low frequency words (Chapter 3) and with cognates vs. non-cognates (Chapter 4). These findings support the hypothesis that lexical processing and, particularly, lexical activation facilitates the acquisition of L2 structures, but that this facilitation is modulated by cross-linguistic syntactic similarity.

3. I extended MOGUL's[1] claims about the representation, processing and acquisition of L1 and L2 lexicon and syntax within the bilingual mind. Specifically, I elaborated on how MOGUL could account for the representation, processing and/or acquisition of cross-linguistically similar and dissimilar L2 structures, words varying in frequency and cognate status and the influence that processing these words would have on the acquisition of L1-L2 similar and dissimilar constructions (Chapters 1 to 4). This constitutes a strong theoretical effort, since to my knowledge no approach to L2 acquisition explicitly covers all these aspects.

[1] Modular On-line Growth and Use of Language framework (Sharwood Smith, 2017; Sharwood Smith & Truscott, 2014).

In Section 1 in this chapter, I review the main findings of this dissertation and their implications, generally outlined in the first two previous paragraphs. In Section 2, I summarise my contributions to the MOGUL framework, briefly presented in the third paragraph. To conclude, in Section 3 I discuss some possible directions for future research.

1. Main findings

In this dissertation, I first investigated whether the similarity between L1 and L2 structures facilitated acquisition by novice L2 learners (Chapter 2). This facilitation is predicted by several approaches to L2 acquisition, which implicitly or explicitly assume that cross-linguistically similar structures are processed using L1 representations that are part of the linguistic system from the start of L2 acquisition, but that cross-linguistically dissimilar structures have to be acquired from input (e.g. Carroll, 1999, 2001; MacWhinney, 2005; B. D. Schwartz & Sprouse, 1994, 1996; Westergaard, 2021). Evidence in favour of this facilitative effect mostly comes from learners with low, intermediate or advanced L2 proficiency, who show more target-like processing and use of cross-linguistically similar (morpho)syntactic structures and features compared to cross-linguistically dissimilar ones (e.g. Bardovi-Harlig, 1997; Chang & Zheng, 2015; C. Ellis et al., 2013; Izquierdo & Collins, 2008; Tokowicz & MacWhinney, 2005). However, the positive influence of cross-linguistic syntactic similarity on L2 syntax acquisition by complete beginner adult learners is under-researched. In Experiment 1, I addressed this gap by having Spanish natives without knowledge of Galician learn Spanish-Galician similar and dissimilar embedded clauses. Learners were incidentally exposed to the structures via an auditory Plausibility Judgement Task. Then, an auditory Grammaticality Judgement Task (GJT) tested learning of the structures. This weakly suggested that the similar structure was part of learners' linguistic system, but provided no evidence that the dissimilar structure had been learnt. I argued that the experimental design was not adequate nor to elicit knowledge of the similar structure nor to produce learning of the dissimilar structure.

To correct these shortcomings, in Experiment 2 I made some changes to the previous experimental design, moving from an implicit to an explicit learning paradigm. I turned the exposure task into a structure-search task, I doubled the number of sentences formed by the similar and the dissimilar structure in the exposure phase, I presented sentences aurally and visually in the exposure and testing phases and I included feedback in the GJT. In this case, the test revealed that the similar and the dissimilar structure were both established in learners' linguistic system and that learning was significantly greater for the similar structure than for the dissimilar one, showing thus a facilitation of cross-linguistic syntactic similarity in L2 learning. I explained this result within MOGUL, arguing that since the similar structure had been repeatedly used in the L1, but the dissimilar structure had only been processed in Experiment 2, the first structure had a higher resting activation level than the second and, hence, was more firmly established in the linguistic system. On the other hand, in Experiment 2 I also addressed two methodological questions. First, I asked whether the dissimilar structure was learnt during the exposure phase or during the test, with the help of feedback.

Learners discriminated between grammatical and ungrammatical stimuli already in the first 20 trials of the GJT, suggesting that the dissimilar structure was learnt during exposure. Second, I investigated the effect of feedback on acquisition. Previous studies have shown that feedback can facilitate L2 syntax learning (e.g. Leeman, 2003; Mackey & Philp, 1998; Muranoi, 2000; Rosa, 1999). In line with this, feedback improved learning of the structures from the first to the last 20 trials of the GJT. Learning of the similar structure was greater than of the dissimilar structure in the first trials of the GJT and feedback did not change this by the end of the test. Finally, a verbal report allowed assessing the conscious or unconscious knowledge of the structures resulting from implicit and explicit learning conditions. Since this awareness measure was included in all the experiments of the dissertation, I summarise the conclusions drawn from all verbal reports at the end of this section.

In Chapters 3 and 4, I focused on whether and how processing words varying in activation due to their frequency of occurrence and cognate status affected the acquisition of L2 structures with or without a similar L1 counterpart. Previous research has not examined the influence of lexical frequency on L2 syntax acquisition, and only a couple of studies have shown that it might influence processing of L1 and L2 structures (Hopp, 2016; Tily et al., 2010). Likewise, the influence of cognates on L2 syntax acquisition has not been investigated and just a few studies have shown that these words might facilitate syntactic processing (X. Chen et al., 2023; Hopp, 2017; J. Huang et al., 2019; Soares et al., 2018, 2019). In Chapters 3 and 4, I hypothesized how differences in lexical frequency and cognateness could influence the acquisition of cross-linguistically similar and dissimilar L2 structures within MOGUL and I provided experimental evidence in support of these hypotheses.

In Chapter 3, I presented Experiment 3, which replicated Experiment 2 but using cognate verbs of a significantly lower frequency in Spanish, participants' L1. First, I hypothesized that the stronger activation of high frequency verbs (Experiment 2) compared to low frequency verbs (Experiment 3) would facilitate the acquisition of the cross-linguistically dissimilar L2 structure. This is because, based on MOGUL, I proposed that the higher the activation of a word is, the higher the current activation level of the structure containing it is during processing, the higher the resting activation level of the structure is when processing terminates and, hence, the more firmly the structure is established in the linguistic system. A comparison of the GJTs in Experiments 2 and 3 confirmed this hypothesis, since it showed that sensitivity to the dissimilar structure was significantly higher when learnt with high frequency verbs than with low frequency verbs. Second, I hypothesized that the facilitative effect of lexical frequency would not obtain for the cross-linguistically similar L2 structure; this would be processed using an L1 structure established in the linguistic system with a high resting activation level and, hence, would be less affected by differences in lexical frequency. This hypothesis was also retained, since sensitivity to the similar structure did not differ between Experiments 2 and 3. Finally, I hypothesized that the learning advantage for the similar structure over the dissimilar one observed in Experiment 2 would replicate in Experiment 3. However, I expected this advantage to be larger when the structures were learnt with low frequency verbs than with high frequency verbs. In support of this hypothesis, sensitivity to the similar structure was greater than to the dissimilar structure in both experiments, but the magnitude of the effect was larger when learning the structures with low frequency verbs.

In Chapter 4, I presented two implicit learning experiments testing whether and how cognates affected the acquisition of cross-linguistically dissimilar L2 structures (Experiment 4) and cross-linguistically similar L2 structures (Experiment 5). For that purpose, I created two versions of a mini-language based on Basque: one with Basque nouns non-cognate with Spanish and Basque verbs cognate with Spanish (*the cognate version*) and the other with Basque nouns and verbs non-cognate with Spanish (*the non-cognate version*). The learning target were verb-final SOV and OSV structures with postpositional agent-patient marking, existing in Basque but not in Spanish (Experiment 4) or verb-medial SVO and OVS structures with prepositional patient marking, similar to Spanish (Experiment 5). In each experiment, two groups of Spanish natives without knowledge of Basque learnt the vocabulary of either the cognate or the non-cognate language version with the help of pictures and the L1 translations. Learning was tested in a picture-word matching task and a picture-naming task performed until 100% accuracy. Previous research has found that adults and children learn cognates more easily than non-cognates (Antón & Duñabeitia, 2020; Comesaña et al., 2019; Marecka et al., 2021; Valente et al., 2018 and more). This learning advantage was also observed in Experiments 4 and 5, since cognate learners reached 100% accuracy in fewer attempts to one or both learning tasks than non-cognate learners did. Simply put, L2 learners mastered cognate verbs faster than non-cognate verbs. Several models of vocabulary learning could explain this effect (e.g. Grainger et al., 2010; Hall, 2002; Kroll & Stewart, 1994). I attributed this result to the cross-linguistic similarity between cognates and their L1 translations, which caused that cognates were established in the linguistic system more easily and more firmly than non-cognates.

Following this vocabulary learning, the two groups of participants were exposed to sentences formed by the target structures and with either a cognate or a non-cognate verb, each accompanied by a picture. Then, they learnt new non-cognate verbs, later used in two tasks testing learning of the structures: a sentence-picture congruency task, which additionally tested vocabulary knowledge, and a written picture-description task. On the one hand, the congruency task yielded no reliable results, since in both experiments some learners consciously judged the congruency of sentence-picture pairs using only their vocabulary knowledge and not their syntax knowledge. On the other hand, the picture-description task showed that learning L2 structures with cognate verbs as opposed to non-cognate verbs led to an overall more accurate use of the structures when these were cross-linguistically dissimilar (Experiment 4), but not when these were cross-linguistically similar (Experiment 5). This suggests that cognates facilitated the acquisition of L1-L2 dissimilar structures, but not of L1-L2 similar structures. Like in Chapter 3, I explained these results within the MOGUL framework. I argued that the stronger activation of cognates compared to non-cognates spread to the L2 structures, resulting in a higher current activation level during processing and a higher resting activation level after processing. This translated into greater learning for the structures processed with cognate verbs compared to the ones processed with non-cognate verbs, but only when these structures could not be processed using a similar L1 counterpart stored in the linguistic system with a high resting activation level.

Finally, I survey the results of the verbal reports conducted at the end of the experiments. L2 syntax learning under implicit (incidental) conditions often results in unconscious or non-verbalizable syntax knowledge (e.g. Kim & Fenn, 2020; Leung & Williams, 2006; Rebuschat, 2009; Tagarelli et al., 2016; Williams, 2005). Conversely, L2 syntax learning under explicit (intentional) conditions usually produces conscious or verbalizable knowledge (N. C. Ellis, 1993; Rebuschat, 2009; Robinson, 1997; Tagarelli et al., 2016). My explicit learning studies (Experiments 2 and 3) matched the results of previous research, for they mostly yielded explicit knowledge of the L2 structures. By contrast, excluding Experiment 1, for which no clear learning of the structures was found, my implicit learning studies (Experiments 4 and 5) also resulted in explicit knowledge of L2 syntax for a large number of participants. I discussed that this result could be a consequence of the exposure task, which did not instruct learners to focus on the form of the input but neither directed attention away from it, and of adult learners having metalinguistic awareness, which could have led them to search for regularities in the input.

2. Contributions to a framework to study L2 acquisition

In Chapter 1 in this dissertation, I mentioned that investigating the facilitative role of cross-linguistic syntactic similarity and processing of words varying in frequency and cognateness in initial L2 syntax acquisition required understanding language processing and acquisition in the bilingual mind. Specifically, I stressed the importance of comprehending how cross-linguistically similar and dissimilar L2 structures were represented, processed and acquired, how high and low frequency words, cognates and non-cognates were represented and processed and how lexical and syntactic processing interacted during L2 acquisition. As detailed in Chapter 2, several theories and models of L2 acquisition can account for how the L2 learner represents, processes and acquires cross-linguistically similar and dissimilar structures (Carroll, 1999, 2001; MacWhinney, 2005; B. D. Schwartz & Sprouse, 1994, 1996; Westergaard, 2021). In spite of their differences, these approaches all predict that L1 syntax will positively influence the acquisition of similar L2 structures (see also Section 1). In addition, as discussed in Chapters 3 and 4, several proposals have been put forth regarding the representation and processing of high and low frequency words, cognates and non-cognates, the most influential one arguably within the BIA+ model of word recognition and comprehension (Dijkstra & van Heuven, 2002). However, hypothesizing how processing these lexical items would influence the acquisition of cross-linguistically similar and dissimilar L2 structures was more challenging. Lexical processing and syntactic processing have traditionally constituted separate research fields and how the two interact during L2 syntax processing and acquisition was not obvious.

In Chapter 3, I mentioned that models of lexical processing accounting for processing of high and low frequency words, such as the BIA+, do not address how words are processed as part of a syntactic structure nor how this syntactic structure may affect lexical processing. In addition, I pointed out that the most influential models of sentence processing (garden-path models, e.g. Frazier, 1987, 1989; Frazier & Clifton, 1997 and constrained-based lexicalist models, e.g. MacDonald et al., 1994; McClelland et al., 1989; Trueswell & Tanenhaus, 1994) do not discuss how differences in lexical frequency may affect sentence processing. This made it quite challenging to provide an explanation for the finding that lexical frequency might facilitate L1 and L2 syntactic processing (Hopp, 2016; Tily et al., 2010). I mentioned that a possible interpretation lied within Hopp's Lexical Bottleneck Hypothesis (Hopp, 2018), according to which slowdowns in lexical processing (e.g. slower processing of low frequency words compared to high frequency words) may lead to deferred or incomplete syntactic processing. Regarding the effect of lexical frequency on syntax acquisition, there was evidence that, in line with usage-based accounts of L1 and L2 acquisition, repeated exposure to a syntactic construction with the same word eased acquisition and generalization of that construction (e.g. Casenhiser & Goldberg, 2005; Goldberg et al., 2004, 2007; Mcdonough & Kim, 2009; Nakamura, 2012). Yet, to my knowledge, no study had investigated whether processing (cross-linguistically similar and/or dissimilar) structures with words having different frequencies in a natural language influenced acquisition of these structures. I knew of no model or theory of L2 acquisition explicitly hypothesizing whether and how this would occur.

In Chapter 4, I commented on several studies showing that cognates might ease syntactic processing. Specifically, I reviewed evidence that embedding cognates in cross-linguistically similar structures might increase their availability for production and comprehension (Cai et al., 2011; X. Chen et al., 2023). I addressed how this was explained by the Shared Syntax account (Hartsuiker et al., 2004; Hartsuiker & Bernolet, 2017; Hartsuiker & Pickering, 2008), which attributes this facilitation to a stronger residual activation of the combinatorial node containing syntactic information after processing a sentence with a cognate than with a non-cognate (see J. Huang et al., 2019). I also summarised some studies indicating that when a sentence is (temporarily) ambiguous between an L1 and an L2 parse, cognates might help achieve a successful L2 computation (Hopp, 2017; Soares et al., 2018, 2019). This finding can be explained by the Lexical Bottleneck Hypothesis, which, as mentioned, claims that a less costly lexical processing (as when processing cognates, compared to non-cognates) can favour a target-like syntactic processing. Nevertheless, I commented that, as far as I knew, no study had investigated whether cognates influenced the acquisition of (L1-L2 similar and dissimilar) syntactic structures, and no model or theory of L2 acquisition expressly addressed this.

Throughout this dissertation, I have presented MOGUL as a theoretical framework that provides a comprehensive overview of L1 and L2 representation, processing, acquisition and interaction in the bilingual mind, and which is detailed enough to allow extrapolating its claims to derive hypotheses about the specific topics studied in this thesis. Accordingly, in Chapters

1 to 4 I have proposed how the MOGUL framework could account for the representation, processing and acquisition of cross-linguistically similar and dissimilar L2 structures, the representation and processing of high frequency and low frequency words, cognates and non-cognates and the influence that processing these words has on the acquisition of L1-L2 similar and dissimilar structures. In what follows, I summarise MOGUL's claims regarding each of these aspects and my contributions to the framework.

2.1. Representation, processing and acquisition of cross-linguistically similar and dissimilar L2 structures

MOGUL framework

The MOGUL framework does not explicitly mention how L1-L2 similar and dissimilar structures are represented, processed and acquired. Yet, I argued that this could be inferred from the way it generally discusses the initial state of L2 acquisition, L2 acquisition by processing and the influence of the L1 on L2 development. The initial state of L2 acquisition includes the processors and information stores innate to the linguistic system. The latter contain all L1 features acquired by the time that L2 development starts and these have high resting activation levels due to their previous use in the L1 (Sharwood Smith & Truscott, 2014, sec. 10.3). Sharwood Smith and Truscott claim that processing is intrinsically cross-linguistic; relevant items from all the languages that a speaker knows activate and compete to be included in the representation of the input irrespective of the language being processed (Sharwood Smith & Truscott, 2014, sec. 6.5.2). This is particularly relevant in the context of MOGUL's *Acquisition by Processing Theory,* which conceives acquisition as simply the result of processing. As a matter of example, the authors propose that when processing a novel L2 word, its phonological form activates and has to be coindexed with an item in the syntactic store. Since existing L1 items are available and have high resting activation levels, the syntactic processor will select one of these items and it will be coindexed with the phonological form of the L2 word. A new representation will not be constructed unless the existing representations are problematic. In line with this, Sharwood Smith and Truscott comment that whenever a processor necessitates establishing a new representation for the input, it does, just as a processing mechanism. Any new representation initially has a very low resting activation level. If it is used in subsequent processing, its resting activation level increases and it gradually becomes a stable item in the linguistic system, i.e. it is acquired (Sharwood Smith & Truscott, 2014, sec. 4.2, 7.4.2).

My contribution

First, from MOGUL's conceptualization of the initial state of L2 acquisition, I inferred that L1 syntactic structures would be present when L2 development began and that they would be established in the syntactic store with a high resting activation level. Second, considering MOGUL's claims about the cross-linguistic nature of processing, I deduced that L1 syntactic structures would be available for use during L2 processing. Finally, extrapolating MOGUL's general view of acquisition by processing to acquisition by processing of L2 syntactic structures, I proposed that when an L2 structure was first encountered, the syntactic processor would attempt to process it using a compatible L1 structure. The construction of a new syntactic representation would not occur unless using one of the existing representations proved inadequate or impossible. Consequently, I hypothesized that if an L2 structure could be processed using a cross-linguistically similar syntactic representation this would occur, and no new representation would be constructed. Conversely, if an L2 structure could not be successfully processed using L1 representations, as it would happen for cross-linguistically dissimilar L2 structures, the syntactic processor would create an appropriate representation. At first, the resting activation level of this new representation would be very low. Then, it would gradually increase as a result of processing, ultimately causing that syntactic structure to become firmly established in the linguistic system.

2.2. Representation and processing of high frequency and low frequency words

MOGUL framework

The MOGUL framework does talk about frequency of occurrence and its role in processing. In short, Sharwood Smith and Truscott relate the notion of frequency to that of resting activation level, assuming that every time that an item is used in processing its resting activation level increases (Sharwood Smith & Truscott, 2014, sec. 4.6.5). The authors maintain that the increase in resting activation level derived from processing occurs for all items in the lexical stores and, hence, do not focus their discussion about frequency on high vs. low frequency words. However, they do mention that MOGUL's resting activation level is exactly like the notion in lexical access research, which assumes that the frequency with which a lexical item is encountered is coded in that item's resting activation level and that this, in turn, determines the speed of lexical access (Sharwood Smith & Truscott, 2014, sec. 4.6.5). Likewise, in MOGUL the resting activation level determines how quickly items are available for processing (Sharwood Smith & Truscott, 2014, sec. 3.3.5). Sharwood Smith and Truscott also incorporate into their framework the logarithmic relation between frequency and resting activation level from lexical processing research. This is arguably responsible for the finding that at really high frequency levels, the difference between two words, even if large, does not affect lexical access (Sharwood Smith & Truscott, 2014, sec. 4.6.5). Finally, MOGUL also associates the notion of resting activation level to that of current activation level, defined as the sum of an item's resting level and the activation obtained during processing, and determining which representations from all the active ones are selected for processing (Sharwood Smith & Truscott, 2014, sec. 3.3.5, 3.4.1).

My contribution

Considering the claims above, I assumed that MOGUL supported a representation of high frequency and low frequency words where the former had a higher resting activation level and, thus, were processed faster than the latter. I argued that, in this way, MOGUL could account for previous evidence that high frequency words are processed faster than low frequency words (e.g. de Groot et al., 2002; Duyck et al., 2008; Gollan et al., 2011; Monsell et al., 1989). Additionally, given the relation between an item's resting activation level and its current activation level, I hypothesized that the higher resting activation level of high frequency words compared to low frequency words would cause them to have a higher current activation level as well. Considering that in MOGUL, the representation with the highest current activation level is usually the one selected for processing, I proposed that the higher current activation level of high frequency words compared to low frequency words should cause that the former were most often selected for processing than the latter. I argued that this would explain why speakers recognize and process high frequency words more accurately than low frequency words (e.g. Duyck et al., 2008; Gollan et al., 2011).

2.3. Representation and processing of cognates and non-cognates

MOGUL framework

The MOGUL framework does not address how cognates are represented and processed. However, I argued that this could be extrapolated from the way it generally describes word representation and processing. In brief, MOGUL conceives words as chains of coindexed acoustic, orthographic, phonological, syntactic and conceptual representations, each in its own module (Sharwood Smith & Truscott, 2014, sec. 2.3.3). Cognates and non-cognates are translation equivalents and, thus, have similar meanings across languages. In this respect, Sharwood Smith and Truscott mention that when an L2 word is learnt, it is usually associated with the conceptual representation of its L1 translation equivalent and that this only changes if suggested by the context (Sharwood Smith & Truscott, 2014, sec. 7.6.1). During processing, representations that share features with the input in any language that the speaker knows activate in each lexical store. Activation spreads bidirectionally to coindexed representations in adjacent modules. Active representations in each module compete for selection and the representation with the highest current activation level wins the competition. Representations with a higher current activation level land at a higher resting activation level after processing and the higher the resting activation level of a representation is, the faster it is available for subsequent processing (Sharwood Smith & Truscott, 2014, sec. 3.3.5, 3.4.1).

My contribution

First, given that MOGUL assumes that activation spreads within lexical stores based on similarity with the input, I assumed that since cognates share orthographic and/or phonological features across languages, when processing these words two chains of representations would activate: that of the cognate in the target language and that of its equivalent in the non-target language. By contrast, since non-cognates share no formal features across languages, when processing these words only the chain of representations of the word in the target language would activate. In addition, considering how Sharwood Smith and Truscott describe the establishment of meaning for L2 words, I inferred that translation equivalents shared a conceptual representation across languages. I assumed that this shared conceptual representation would be more strongly activated for cognates than for non-cognates, for in the first case it would receive activation from two chains of representations, but in the second case, it would receive activation from just one. Since in MOGUL activation is bidirectional, I also assumed that the stronger activation of cognates' conceptual representation would spread back to the chain of representations of the target word, causing that, overall, cognates had a higher current activation level than non-cognates. Finally, considering the relation between current activation level and resting activation level, I hypothesized that if cognates had a higher current activation level than non-cognates, they would also have a higher resting activation level when processing terminates. The higher resting and current activation level of cognates compared to non-cognates would cause that the former were processed faster and more accurately than the latter, as evidenced in the literature (e.g. Dijkstra et al., 1999; Lemhöfer & Dijkstra, 2004; Van Assche et al., 2009, 2011).

2.4. The effect of lexical processing on the acquisition of cross-linguistically similar and dissimilar L2 structures

MOGUL framework

MOGUL does not address how lexical processing, including processing of words differing in frequency or cognate status, would affect acquisition by processing of cross-linguistically similar or dissimilar L2 structures. I hypothesized how MOGUL could account for this based on how I claimed that high and low frequency words, cognates and non-cognates, and cross-linguistically similar and dissimilar L2 structures would be processed in the framework (Sections 2.1-2.3) and on how it describes the relation between lexical and syntactic processing. When discussing the acquisition by processing of syntactic constructions, Sharwood Smith and Truscott describe how processing such constructions necessarily involves processing the words that form them and how processing at the lexical level interacts with processing at the construction level. In brief, the authors propose that processing in the syntactic module involves two roughly simultaneous sources of activity. On the one hand, as words in the input are incrementally perceived, their syntactic representations activate and the syntax processor combines them into a larger syntactic representation. The way in which the syntax processor constructs this syntactic representation is constrained by Universal Grammar (UG), in line with generative approaches of language development (e.g. Chomsky, 1995). On the other hand, as the syntactic representations of words activate, activation spreads to larger syntactic representations containing them, possibly including that of the target construction. This process raises the current activation level of these larger representations, which try to impose themselves on subsequent input. This source of activity, which involves activating representations of constructions as a whole instead of building them from their components, is compatible with Construction Grammar (Sharwood Smith & Truscott, 2014, sec. 4.5.3). Crucially, the current activation level of a representation influences the degree of the rise in the current activation level of other representations sharing features with it in the same store. In addition, as mentioned in the previous sections the current activation level of an item determines the extent of the increase in its resting activation level after processing (Sharwood Smith & Truscott, 2014, sec. 3.3.5). However, the increases in resting activation level derived from processing gradually diminish as the resting activation level becomes higher. This has consequences for acquisition by processing, since the strength of an item's resting activation level determines whether it is a more or less stable item in the linguistic system (Sharwood Smith & Truscott, 2014, sec. 4.2).

My contribution

From MOGUL's assumption that a word's syntactic representation activates larger syntactic representations including it and that a representation's current activation level determines the rise in current activation level of other representations with common features, I inferred that the higher the current activation of a word was, the larger the increase in current activation of the structure containing it was. Considering this, I hypothesized that the higher current activation level of high frequency words and cognates, compared to low frequency words and non-cognates, would cause that L2 structures processed with the first pair of words had a higher current activation level than those processed with the second pair of words. In addition, since a higher current activation level results in a higher resting activation level, I argued that L2 structures would also have a higher resting activation level when processed with high frequency words or cognates than with low frequency words or non-cognates. On the other hand, I hypothesized that cross-linguistically similar L2 structures were stored in the linguistic system with a high resting activation level, but that cross-linguistically dissimilar ones initially received a low resting activation level (Section 2.1). Given that the higher the resting activation level of a representation is, the smaller the increases in this resting level derived from processing are, I assumed that processing L1-L2 similar structures would result in smaller increases in resting activation level than processing L1-L2 dissimilar structures. Consequently, I argued that the difference in resting activation level after processing sentences with high vs. low frequency words or cognates vs. non-cognates would be small or non-existent for cross-linguistically similar structures, but that it would be larger for cross-linguistically dissimilar structures. Since MOGUL relates an item's resting activation level to its stability in the linguistic system, I concluded that cross-linguistically similar structures would be comparably established in the linguistic system regardless of the frequency and the cognateness of the words they included. By contrast, cross-linguistically dissimilar structures would be more robustly established in the system when processed with high frequency words and cognates than with low frequency words and non-cognates.

2.5. Other contributions

In the previous sections, I have presented my major contributions to the MOGUL framework. Nevertheless, throughout the dissertation, I also extrapolated MOGUL's claims to account for other aspects relevant for my experiments, such as the role of feedback in learning L2 structures and the cognate advantage in L2 word learning.

MOGUL framework and feedback

Sharwood Smith and Truscott briefly comment on the effect of feedback on language learning. They mention that corrective feedback is neither necessary nor particularly useful for L1 acquisition to take place, but that in some cases older children and L2 learners do respond to feedback (Sharwood Smith & Truscott, 2014, sec. 1.7.1). For instance, they note that negative feedback can facilitate the acquisition of L1 and L2 word meaning (Sharwood Smith & Truscott, 2014, sec. 5.4.3, 7.6.1). The authors do not explicitly address how feedback could affect the acquisition of (L1 or L2) syntactic structures. Nevertheless, I argue that this can be inferred from the assumption that if a representation proves appropriate for processing, its resting activation level increases, but if it is not useful for future parsing, its resting activation level does not increase and it does not become part of the syntax (Sharwood Smith & Truscott, 2014, sec. 4.3).

My contribution

In Chapter 2, I discussed how the feedback that participants received during the GJT of Experiment 2 revealed whether a structure was correct and, hence, useful for future processing. In this light, I proposed that feedback could cause that the resting activation level of the cross-linguistically similar and dissimilar L2 structures increased, but that the resting activation level of their ungrammatical counterparts did not. This would cause that the first constructions were firmly established in the linguistic system, but that the second did not become part of the grammar.

MOGUL framework and learning of cognates

As mentioned in Section 2.3, MOGUL does not address cognate representation and processing. Likewise, it does not discuss cognate learning. Regarding word learning more generally, MOGUL only accounts for the development of L1 and L2 word meaning (Sharwood Smith & Truscott, 2014, sec. 5.4.3, 7.6.1). Focusing on the L2, Sharwood Smith and Truscott comment that when an L2 word is learnt together with its L1 translation, the chain of representations of the L1 word activates. Similarly, the orthographic/phonological and syntactic representations of the novel L2 word activate and must be coindexed with a conceptual representation. This will most likely be the representation of its L1 counterpart, for it will be highly active. If the meaning of the L2 word is not significantly different from the meaning of the L1 word, the L2 item will be processed and used simply as a translation of the L1 item.

My contribution

Based on the ideas above and on how I hypothesized that MOGUL could account for cognate and non-cognate representation and processing, I postulated some hypotheses regarding how, in Chapter 4, participants learnt cognate and non-cognate verbs presented with their L1 translations. In short, I argued that due to the cross-linguistic activation of formally similar words, the chains of representations of L1 verbs were more strongly activated when they were cognate with their L2 equivalents than when they were non-cognate with their L2 counterparts. I argued that the stronger activation of the syntactic and conceptual representations of the former compared to the latter made it easier to coindex these representations with the orthographic and phonological representations of L2 cognates than with those of L2 non-cognates. Consequently, I claimed that cognates were learnt more easily than non-cognates.

3. Directions for further research

The experiments in this dissertation have addressed two gaps in the literature on adult L2 syntax acquisition. On the one hand, the positive influence of cross-linguistic syntactic similarity on the initial acquisition of syntactic structures, from adults' first exposure to the L2, is under-studied. The experiments presented in Chapters 2 and 3 broaden our understanding of this topic. On the other hand, the facilitative effect of high frequency words and cognates on the initial acquisition of cross-linguistically similar or dissimilar L2 structures had not been investigated. The experiments in Chapters 3 and 4 are a first step towards understanding this issue. Further research is needed to replicate the results of this thesis, since reproducing the findings of experimental research is of upmost importance to strengthen the reliability of any scientific claims made (Open Science Collaboration, 2015). In what follows, I discuss three additional directions for future research.

3.1. Assessing the pedagogical implications of the results of the dissertation

This dissertation has investigated how adults' initial L2 syntax acquisition can be facilitated, a topic that is highly relevant in an era of globalization in which learning languages is increasingly more important (see Chapter 1). Although my work belongs to the field of psycholinguistics, I consider that my findings can interest other fields such as second or foreign language teaching and can promote interdisciplinary research. It might be tempting to extract pedagogical implications from the findings reported in Section 1 and suggest, for instance, that language teachers exploit cross-linguistic syntactic similarity and use cognates and high frequency words to facilitate learning of syntactic structures in the L2 classroom. However, caution is advised, for the findings presented in this thesis are the result of experiments conducted in a highly controlled laboratory setting and whether and how the same findings would obtain in a different learning context is not clear.

There are two main approaches to foreign language learning in the classroom: the teacher-centred approach and the learner-centred approach (Markina & Garcia, 2022; Sánchez, 2007). The teacher-centred approach is the traditional one, where the teacher conveys knowledge to the students usually using explicit metalinguistic explanations (for instance, based on grammars in foreign language textbooks). Learners then do controlled exercises with little room for creativity to consolidate these theoretical explanations. These learning conditions are recreated in explicit L2 learning experiments where learners are instructed one or more grammatical rules and then practice using these rules in one or various language tasks (e.g. Bastarrika & Davidson, 2017; Tagarelli et al., 2016; Tolentino & Tokowicz, 2014). On the other hand, in the learner-centred approach learners are not taught language rules explicitly, but have to induce them from their experience with the L2. This approach to L2 learning is closer to the one taken in this thesis. As in the learner-centred approach, in the experiments I conducted learners were not taught the grammatical structures of the mini-languages. Instead, they had to extract regularities from input on their own. However, the way L2 learning is elicited in the classroom is different from the way I did this in my experiments. In learner-centred teaching, learners are encouraged to actively analyse and discuss L2 input, both with the teacher and with other students, and are asked to use the L2 in writing and in speaking as much as possible. By contrast, in the experiments in this thesis participants learnt the structures by being exposed to the L2 individually. In some experiments they were asked to actively think or analyse the input to discover the target structures (Experiments 2 and 3), but in the rest of experiments they were not. In none of the experiments did they interact with other participants. In addition, learners were tested on their knowledge, also individually, by means of highly controlled tests which did not require writing or talking, or which required this but allowed for almost no creativity (Experiments 4 and 5).

In sum, the teaching techniques and the tools used in the L2 classroom differ from those used in my psycholinguistic experiments. To assess the real pedagogical implications of the findings of this thesis, it would be necessary to design experiments that recreate the learning conditions in classroom settings to test whether the facilitative role of cross-linguistic syntactic similarity and lexical processing observed also obtains in these circumstances.

3.2. Exploring the facilitative effect of cross-linguistic syntactic similarity and lexical processing beyond the earliest stage of L2 syntax acquisition

This thesis studied the acquisition of syntactic structures at the earliest stage of L2 development. All the experiments were conducted with adults who had never been exposed to the L2s on which the mini-languages were based and their first encounter with these languages was in the lab. Since my focus was on the initial stages of learning, I conducted short experiments, performed in one session and testing learning after a relatively brief exposure to the L2 structures (in all experiments, the exposure phase lasted for 10 minutes or less and the total duration of the experiments ranged from 45 minutes to 1 hour and 30 minutes). Further research could explore whether and how the facilitation exerted by cross-linguistic syntactic similarity, high frequency words and cognates in the acquisition of L2 structures varies when experience with the structures increases, as it would naturally occur in the course of L2 development. To this aim, future studies could replicate the experiments in this thesis but adding a second experimental session conducted on a different day. In this second session, participants would receive further exposure to the L2 structures and would be tested again on their syntax knowledge. Based on how the MOGUL framework proposes that syntactic representations develop within the bilingual mind, I hypothesize that the facilitation stemming from cross-linguistic syntactic similarity and lexical processing would gradually disappear as learners achieve greater proficiency in the L2.

On the one hand, the learning advantage of the cross-linguistically similar L2 structure over the cross-linguistically dissimilar one (Chapters 2 and 3) was attributed to the fact that the former could be processed using an L1 syntactic representation firmly stored in the linguistic system with a high resting activation level. By contrast, the latter could not be processed using an L1 syntactic representation and an appropriate one had to be constructed. This representation initially had a low resting activation level and had to gradually become consolidated in the linguistic system (see Section 2.1). As mentioned, MOGUL assumes that each time that an item is used in processing its resting activation level increases and it becomes a more stable element. Yet, processing an item will not increase its resting activation level indefinitely. At some point, the resting level will be so high that it will rise no more. I hypothesize that, during the second experimental session, the resting activation level of the similar and the dissimilar structure will continue to increase. As the structures become more consolidated in the linguistic system, the learning difference between the two may strengthen. However, if the resting activation level of the similar structure is higher than that of the dissimilar structure, as suggested in Chapters 2 and 3, it is possible that at some point during this second session, the level of the similar structure ceases to increase. Consequently, the difference in resting activation level between the similar and the dissimilar structure would narrow down, and the two structures could become comparably established in learners' linguistic system. In familiar thinking terms, the learning advantage of the similar structure over the dissimilar one in Chapters 2 and 3 could not replicate. On the other hand, the finding that processing cross-linguistically dissimilar L2 structures with high frequency words and cognates as opposed to low frequency words and non-cognates facilitated their acquisition (Chapters 3 and 4) was also attributed to a difference in resting activation level between the structures processed with the two pairs of words (see Section 2.4). I hypothesize that if learners processed more instances of the dissimilar structures in a second experimental session, the resting activation level of these constructions could raise to the point that any advantage in learning resulting from processing them with high frequency words or cognates vs. low frequency words or non-cognates could become non-significant.

3.3. Further investigating the facilitative effect of lexical activation on the acquisition of cross-linguistically dissimilar L2 structures

One of the main claims of this dissertation is that lexical activation might facilitate the acquisition of L1-L2 dissimilar structures. Future research should be conducted to gather more empirical evidence supporting this claim. One way to do this would be to replicate Experiment 4 but using a mini-language that included identical cognates (i.e. cognates with exactly the same orthography across languages, e.g. Spanish-Catalan *pintar-pintar,* "paint"), instead of non-identical cognates (i.e. cognates slightly differing in orthography and/or phonology across languages, e.g. Spanish-Basque *pintar-pintatu*). Bilinguals recognise and read cognates with higher orthographic (and/or phonological) overlap faster than cognates with lower overlap (e.g. Dijkstra et al., 2010; Duyck et al., 2007; Van Assche et al., 2011). This could be explained by claiming that as the similarity of the cognate across languages increases, the degree of its cross-linguistic activation also increases and this leads to a larger facilitation in processing. Accordingly, the largest co-activation and, hence, the largest facilitation effect, is found for identical cognates, having a complete form overlap with its translation equivalent in another language (Dijkstra et al., 2010). If, as hypothesized in this thesis, the stronger activation of (non-identical) cognates compared to non-cognates facilitated learning of L2 structures in Experiment 4, then the extent of this facilitation could increase if the study were replicated with identical cognates, having a stronger activation than non-identical cognates.

Finally, another option would be to replicate Experiments 2 and 3, conducted with high frequency and low frequency verbs, respectively, but using a more extreme frequency manipulation. In Experiment 2, there was a quite large variability in the Spanish frequency of the cognate verbs used (exposure verbs, $M = 98.97$, $SD = 106.16$; test verbs, $M = 115.09$, $SD = 112.98$). In Experiment 3, which was designed to be compared with Experiment 2, I selected cognate verbs that had a lower frequency of occurrence in Spanish than the lowest frequency verb in Experiment 2. There was also some variability in the frequency of the verbs used in this experiment, but this was much lower than in Experiment 2 (exposure verbs, $M = 4.76$, $SD = 2.02$; test verbs, $M = 5.13$, $SD = 2.55$). In sum, even if overall mean frequency was significantly higher for high frequency verbs than for low frequency verbs, in some individual cases the difference between a high frequency verb and a low frequency verb was not large. It is often the case that studies looking into word frequency effects in processing use high frequency and/or low frequency stimuli with high variability (e.g. in Lehtonen et al., 2012, stimuli lay within a high frequency range of 7.89–504 per million and a low frequency range of 0.04–4.23 per million, see also Gollan et al., 2008). However, if as hypothesized in Chapter 3, the stronger activation of high frequency verbs compared to low frequency verbs facilitated the acquisition of the cross-linguistically dissimilar L2 structure, then it is possible that a stronger facilitation occurred if differences in frequency and, hence, in activation between the two groups of verbs were even more pronounced. To test this hypothesis, future experiments with different materials could replicate Experiments 2 and 3 but establishing very high and low frequency ranges, with smaller variability than in Experiments 2 and 3, within which high and low frequency verbs were selected.

Appendix A

Appendices to Chapter 2

1. Linguistic questionnaire used in Experiments 1-3

English translation of the questionnaire (original in Spanish)

Personal information		
Name		
Surname		
Sex	Male	Female
Date of birth (year)		
Email		
Phone number		
Place of residence (city and region)		
Place of birth (city and region)		
Have you ever lived in another city and/or region?	Where?	When?

Native language
When you were little, which language did you use with…
…your mother?
…your father?
…your siblings?
…your grandparents?

Age of acquisition
How old were you when you started…
…speaking in Spanish?
…speaking in Basque?
If you have a certificate in Basque, name which:

Language use			
Which language and how often did you use it…			
…when you were little, before starting school?			
	At school/university/work	At home	Other places
…in your childhood, at primary school?			
… in your puberty, at high school?			
… nowadays, as an adult?			
Which language do you feel most comfortable using?	Spanish	Basque	Both

Proficiency (self-assessment)					
Rank your skills in the following languages:					
	Speaking	Listening	Reading	Writing	Certificate (if any)
Basque					
Spanish					
English					
French					
Other					

Response options and scoring

Native language	
Spanish only	1
Mostly Spanish, rarely Basque	2
Mostly Spanish, but Basque at least 25% of the time	3
Spanish and Basque with equal frequency	4
Mostly Basque, but Spanish at least 25% of the time	5
Mostly Basque, rarely Spanish	6
Basque only	7

Language use	
Spanish only	1
Mostly Spanish, rarely Basque	2
Mostly Spanish, but Basque at least 25% of the time	3
Spanish and Basque with equal frequency	4
Mostly Basque, but Spanish at least 25% of the time	5
Mostly Basque, rarely Spanish	6
Basque only	7

Proficiency (self-assessment)	
Very poor	1
Poor	2
Enough	3
Pretty good	4
Good	5
Very good	6
Perfect	7

2. Linguistic information about the participants in Experiments 1 and 2

The tables below report the information obtained in the linguistic background questionnaire that participants in Experiments 1 and 2 filled out before the experiment. For each experiment, I report first the information about learners' language use in different life periods and in different environments. Then, I report learners' self-assessed proficiency in Spanish.

Experiment 1 (n = 24)

	Childhood	**Puberty**	**Adulthood**
School/university/work	2.17 (1.37)	2.08 (1.28)	1.42 (0.83)
Home	1.08 (0.28)	1.04 (0.20)	1.08 (0.28)
Other places	1.13 (0.34)	1.21 (0.42)	1.38 (0.58)
Mean	1.46 (0.96)	1.44 (0.90)	1.29 (0.62)

TABLE A-2.1. Language use (SD in brackets) during childhood, puberty and adulthood in different environments as self-assessed by participants in Experiment 1. Scores are on a 7-point scale: 1 = Spanish only; 2 = Mostly Spanish, rarely Basque; 3 = Mostly Spanish, but Basque at least 25% of the time; 4 = Spanish and Basque with equal frequency; 5 = Mostly Basque, but Spanish at least 25% of the time; 6 = Mostly Basque, rarely Spanish; 7 = Basque only.

Speaking	6.88 (0.45)
Listening	6.96 (0.20)
Reading	6.92 (0.28)
Writing	6.75 (0.53)
Mean	6.88 (0.39)

TABLE A-2.2. Proficiency level speaking, listening, reading and writing in Spanish (SD in brackets) as self-assessed by participants in Experiment 1. Scores are on a 7-point scale: 1 = Very poor; 2 = Poor; 3 = Enough; 4 = Pretty good; 5 = Good; 6 = Very good; 7 = Perfect.

Experiment 2 (n = 44)

	Childhood	**Puberty**	**Adulthood**
School/university/work	2.11 (1.47)	2.11 (1.50)	1.30 (0.59)
Home	1.07 (0.33)	1.05 (0.21)	1.02 (0.15)
Other places	1.16 (0.37)	1.16 (0.37)	1.32 (0.56)
Mean	1.45 (1.01)	1.44 (1.01)	1.21 (0.49)

TABLE A-2.3. Language use (SD in brackets) during childhood, puberty and adulthood in different environments as self-assessed by participants in Experiment 2. Scores are on a 7-point scale: 1 = Spanish only; 2 = Mostly Spanish, rarely Basque; 3 = Mostly Spanish, but Basque at least 25% of the time; 4 = Spanish and Basque with equal frequency; 5 = Mostly Basque, but Spanish at least 25% of the time; 6 = Mostly Basque, rarely Spanish; 7 = Basque only.

Speaking	6.70 (0.59)
Listening	6.84 (0.37)
Reading	6.82 (0.45)
Writing	6.67 (0.64)
Mean	6.76 (0.53)

TABLE A-2.4. Proficiency level speaking, listening, reading and writing in Spanish (SD in brackets) as self-assessed by participants in Experiment 2. Scores are on a 7-point scale: 1 = Very poor; 2 = Poor; 3 = Enough; 4 = Pretty good; 5 = Good; 6 = Very good; 7 = Perfect.

3. Informed consent used in Experiments 1 and 2

This is the Spanish "Informed consent in comprehension tests" participants read and signed before Experiments 1 and 2. This consent provided participants with all the necessary information about the experiment so that they could decide freely and voluntarily whether they wanted to participate. The information given included: the project the study was part of, details of the Principal Investigators of the project and the person in charge of the experiment, description, aims and procedure of the study, risks and rights of the participant and policy of conservation and processing of personal data.

CONSENTIMIENTO INFORMADO EN PRUEBAS DE COMPRENSIÓN

El presente informe tiene como objetivo primordial proporcionarle toda la información necesaria sobre el experimento en el que va a participar y sobre la conservación y tratamiento de sus datos personales, con el objetivo de que pueda decidir libre y voluntariamente sobre su participación en el mismo.

Identificación del proyecto

Título del proyecto: Cross-linguistic activation effects in bilingual language processing and learning
Financiación: Ministerio de Ciencia, Innovación y Universidades
Título del estudio: The effect of syntactic co-activation in L2 syntax learning
Código del proyecto: PGC2018-097970-B-100

Identificación del investigador principal y forma de contacto

Nombre y apellidos: Kepa Erdozia y Mikel Santesteban
Dirección: Centro de Investigación Micaela Portilla 3.2. Dept. Lingüística y Estudios Vascos. Facultad de Letras, Universidad del País Vasco (UPV/EHU)
E-mail: kepa.erdozia@ehu.eus eta mikel.santesteban@ehu.eus
Teléfono: 945013650

Identificación del investigador responsable

El investigador responsable se encargará de pasar la prueba experimental y de informarle adecuadamente.
Nombre y apellidos: Noèlia Sanahuja Cobacho
Dirección: Centro de Investigación Micaela Portilla 3.2. Dept. Lingüística y Estudios Vascos (UPV/EHU)
E-mail: noelia.sanahuja@ehu.eus

DESCRIPCIÓN Y OBJETIVOS DE LA INVESTIGACIÓN

Nuestro objetivo principal es llegar a entender cómo la activación interlingüística de las dos lenguas modula el aprendizaje y el procesamiento de la L2 en el hablante bilingüe. Dentro de este objetivo general trabajaremos con la hipótesis de que la activación interlingüística guía la mayoría de los procesos cuando utilizamos una segunda lengua. Investigaremos cual es el papel de la distancia tipológica entre L1 y L2 en los efectos de interferencia interlingüística durante el aprendizaje y procesamiento de lenguaje y el papel del conocimiento sintáctico en el aprendizaje de reglas de L2.

PROPÓSITO DEL ESTUDIO
El objetivo del presente estudio es investigar cómo el cerebro procesa el lenguaje durante el aprendizaje de una segunda lengua.

PROCEDIMIENTO
El procedimiento a seguir consiste en escuchar y leer unas oraciones mientras se mira la pantalla de un ordenador. Seguidamente, se realizará una prueba lingüística basada en las oraciones que se han escuchado. El experimento se realizará en 1 sesión y tendrá una duración máxima de 75 minutos.

Riesgos e incomodidades
Ninguno de los procedimientos representa peligro alguno para la salud o integridad física. Todas las intervenciones se llevarán a cabo con todas las medidas preventivas requeridas en la situación de Covid-19.

DERECHOS DEL PARTICIPANTE
Cláusula de voluntariedad y derecho de revocación
La información que contienen los datos personales del participante o cualquier otro dato identificativo no se proporcionará a terceros y se protegerá la privacidad de los mismos. Los resultados de este proyecto pueden llegar a publicarse en libros o revistas especializadas o pueden usarse con finalidades didácticas. La participación en este estudio es completamente voluntaria y, como tal, puede revocar el consentimiento dado en cualquier momento, sin dar explicaciones de ningún tipo y sin que ello suponga ningún perjuicio o medida en su contra. De igual forma, a criterio del investigador, usted puede ser retirado del estudio por alguna de las siguientes razones: (a) si no cumple con los requisitos mínimos que se establezcan para participar en el estudio; (b) si por cualquier motivo se interrumpe el estudio.
Cláusula sobre el derecho a tener más información sobre el proyecto
Si colabora en este estudio, una vez haya finalizado, tendrá usted a su disposición toda la información relativa a los resultados obtenidos en el mismo. Para acceder a ella, es necesario que se ponga en contacto con el investigador responsable del proyecto a través de la dirección de e-mail que consta en este documento.

PROTECCIÓN DE DATOS:
Se le informa de que de conformidad al Reglamento Europeo de Protección de Datos (UE2016/679):

- Los datos personales que se le solicitan son:
  a) Datos de carácter identificativo: DNI/NIF, NOMBRE Y APELLIDOS, DIRECCIÓN (POSTAL, ELECTRÓNICA), TELÉFONO, IMAGEN/VOZ
  b) Datos de características personales: FECHA DE NACIMIENTO, LUGAR DE NACIMIENTO, EDAD, SEXO, NACIONALIDAD, LENGUA MATERNA
  c) Datos académicos y profesionales: FORMACIÓN, TITULACIONES
- El código del tratamiento de datos es:TI0091
- El nombre del tratamiento de datos es: DATOS GOGO ELEBIDUNA-MENTE BILINGÜE
- La finalidad de este tratamiento es: CUESTIONARIO DE PERFIL LINGÜÍSTICO DE LOS PARTICIPANTES EN LOS EXPERIMENTOS PSICOLINGÜÍSTICOS DEL GRUPO DE INVESTIGACIÓN "GOGO ELEBIDUNA/MENTE BILINGÜE"
- El responsable del tratamiento de datos es la UPV/EHU:

Identidad: Universidad del País Vasco/Euskal Herriko Unibertsitatea CIF: Q4818001B Dirección postal: Barrio Sarriena s/n, 48940-Leioa (Bizkaia) Página web: www.ehu.eus Datos de contacto del Delegado de Protección de Datos: dpd@ehu.eus

- El periodo de conservación de sus datos será: Los datos se conservarán mientras no se solicite su supresión por la persona interesada y, en cualquier caso, siempre que estén abiertos los plazos de recurso y/o reclamación procedente o mientras sigan respondiendo a la finalidad para la que fueron obtenidos.
- La legitimación del tratamiento es: su consentimiento informado.
- Cesiones y transferencias internacionales de sus datos: No se cederán datos salvo previsión legal. No se efectuarán transferencias internacionales.
- Los derechos sobre sus datos son los de acceso, supresión, rectificación, oposición, limitación del tratamiento, portabilidad y olvido. Puede ejercerlos enviando su petición a dpd@ehu.eus.
- Tiene a su disposición información adicional en http://www.ehu.eus/babestu
- La información completa sobre este tratamiento está en: https://www.ehu.eus/es/web/idazkaritza-nagusia/ikerketa-datu-pertsonalen-tratamenduak

IDENTIFICACIÓN DE LA PERSONA QUE PRESTA EL CONSENTIMIENTO

Yo, ____, con DNI nº __________________ declaro que he leído este documento y que doy mi consentimiento a participar voluntariamente en este estudio.

_______________________________________ _________________

Voluntario/a Fecha

El investigador abajo firmante declara que el participante ha recibido la información escrita y oral necesaria para garantizar que su participación pueda considerarse libre y voluntaria.

_______________________________________ _________________

Investigador/a Fecha

4. List of vocabulary used in Experiments 1 and 2

<u>Exposure set</u>

Impersonal expressions	**Proper nouns**	**Verbs**	**Inanimate nouns (plausible sent.)**	**Inanimate nouns (implausible sent.)** *Only for Experiment 1*
É importante Sp. "es importante" Eng. "it is important"	Pedro	Reparar Sp. "reparar" Eng. "fix"	Radio Sp. "radio" Eng. "radio"	Tornado Sp. "tornado" Eng. "tornado"
	Ángel	Apagar Sp. "apagar" Eng. "turn off"	Consola Sp. "consola" Eng. "console"	Vento Sp. "viento" Eng. "wind"
	Pablo	Acender Sp. "encender" Eng. "turn on"	Móbil Sp. "móvil" Eng. "mobile"	Sol Sp. "sol" Eng. "sun"
	Luis	Vender Sp. "vender" Eng. "sell"	Portátil Sp. "portátil" Eng. "laptop"	Mar Sp. "mar" Eng. "sea"
É posible Sp. "es possible" Eng. "it is possible"	Jorge	Comprar Sp. "comprar" Eng. "buy"	Televisor Sp. "televisor" Eng. "television"	Nube Sp. "nube" Eng. "cloud"
	Alberto	Saborear Sp. "saborear" Eng. "savor"	Torta Sp. "tarta" Eng. "cake"	Río Sp. "río" Eng. "river"
	Adrián	Cortar Sp. "cortar" Eng. "cut"	Carne Sp. "carne" Eng. "meat"	Monte Sp. "monte" Eng. "mount"
	Iván	Cociñar Sp. "cocinar" Eng. "cook"	Pan Sp. "pan" Eng. "bread"	Montaña Sp. "montaña" Eng. "mountain"
É probable Sp. "es probable" Eng. "it is probable"	Raúl	Mastigar Sp. "masticar" Eng. "chew"	Brócoli Sp. "brócoli" Eng. "broccoli"	Estanque Sp. "estanque" Eng. "pond"
	Óscar	Conxelar Sp. "congelar" Eng. "freeze"	Cenoria Sp. "zanahoria" Eng. "carrot"	Galaxia Sp. "galaxia" Eng. "galaxy"
	Mónica	Devolver Sp. "devolver" Eng. "return"	Xoia Sp. "joya" Eng. "jewel"	Universo Sp. "universo" Eng. "universe"
	Isabel	Gardar Sp. "guardar" Eng. "store"	Pulseira Sp. "pulsera" Eng. "bracelet"	Océano Sp. "océano" Eng. "ocean"
É sorprendente Sp. "es sorprendente" Eng. "it is surprising"	Lucía	Encargar Sp. "encargar" Eng. "order"	Abrigo Sp. "abrigo" Eng. "coat"	Amañecer Sp. "amanecer" Eng. "dawn"

Impersonal expressions	**Proper nouns**	**Verbs**	**Inanimate nouns (plausible sent.)**	**Inanimate nouns (implausible sent.)** *Only for Experiment 1*
É necesario Sp. “es necesario” Eng. “it is necessary”	Paula	Envolver Sp. “envolver” Eng. “wrap”	Reloxo Sp. “reloj” Eng. “watch”	Barranco Sp. “barranco” Eng. “gorge”
	Alicia	Enviar Sp. “enviar” Eng. “send”	Xersei Sp. “jersey” Eng. “sweater”	Precipicio Sp. “precipicio” Eng. “cliff”
	Silvia	Gañar Sp. “ganar” Eng. “win”	Concurso Sp. “concurso” Eng. “contest”	Tormenta Sp. “tormenta” Eng. “storm”
	Irene	Perder Sp. “perder” Eng. “lose”	Proba Sp. “prueba” Eng. “test”	Furacán Sp. “huracán” Eng. “hurricane”
	Rosa	Financiar Sp. “financiar” Eng. “finance”	Torneo Sp. “torneo” Eng. “tournament”	Terremoto Sp. “terremoto” Eng. “earthquake”
	Andrea	Organizar Sp. “organizar” Eng. “organize”	Carreira Sp. “carrera” Eng. “race”	Incendio Sp. “incendio” Eng. “fire”
	Carmen	Cancelar Sp. “cancelar” Eng. “cancel”	Campeonato Sp. “campeonato” Eng. “championship”	Inundación Sp. “inundación” Eng. “flood”

TABLE A-4.1. Impersonal expressions, proper nouns, verbs and inanimate nouns in the exposure set of Experiments 1 and 2. Spanish (Sp.) and English (Eng.) translations are provided below each word.

Testing set

Impersonal expressions	**Proper nouns**	**Verbs**	**Inanimate nouns**
É importante Sp. "es importante" Eng. "it is important"	Antonio	Firmar Sp. "firmar" Eng. "sign"	Carta Sp. "carta" Eng. "letter"
	José	Ler Sp. "leer" Eng. "read"	Libro Sp. "libro" Eng. "book"
	Manuel	Recibir Sp. "recibir" Eng. "receive"	Postal Sp. "postal" Eng. "postcard"
	Francisco	Corrixir Sp. "corregir" Eng. "correct"	Correo Sp. "correo" Eng. "mail"
É posible Sp. "es posible" Eng. "it is possible"	Juan	Escribir Sp. "escribir" Eng. "write"	Novela Sp. "novela" Eng. "novel"
	Alejandro	Consultar Sp. "consultar" Eng. "check"	Factura Sp. "factura" Eng. "invoice"
	Javier	Modificar Sp. "modificar" Eng. "modify"	Informe Sp. "informe" Eng. "report"
	Salvador	Redactar Sp. "redactar" Eng. "write"	Proposta Sp. "propuesta" Eng. "proposal"
É probable Sp. "es probable" Eng. "it is probable"	Carlos	Revisar Sp. "revisar" Eng. "revise"	Comunicado Sp. "comunicado" Eng. "statement"
	Miguel	Aceptar Sp. "acceptar" Eng. "accept"	Lei Sp. "ley" Eng. "law"
	Sofía	Pintar Sp. "pintar" Eng. "paint"	Casa Sp. "casa" Eng. "house"
	Ana	Limpar Sp. "limpiar" Eng. "clean"	Habitación Sp. "habitación" Eng. "room"
É sorprendente Sp. "es sorprendente" Eng. "it is surprising"	Laura	Ordenar Sp. "ordenar" Eng. "tidy up"	Cociña Sp. "cocina" Eng. "kitchen"
	Victoria	Recoller Sp. "recoger" Eng. "clean up"	Almacén Sp. "almacén" Eng. "warehouse"

Impersonal expressions	**Proper nouns**	**Verbs**	**Inanimate nouns**
É necesario Sp. "es necesario" Eng. "it is necessary"	Marta	Reformar Sp. "reformar" Eng. "renovate"	Apartamento Sp. "apartamento" Eng. "apartment"
	Elena	Describir Sp. "describir" Eng. "describe"	Paisaxe Sp. "paisaje" Eng. "landscape"
	Sara	Observar Sp. "observar" Eng. "observe"	Cadro Sp. "cuadro" Eng. "picture"
	Ángela	Admirar Sp. "admirar" Eng. "admire"	Escultura Sp. "escultura" Eng. "sculpture"
	Julia	Fotografar Sp. "fotografiar" Eng. "photograph"	Xardín Sp. "jardín" Eng. "garden"
	Alba	Contemplar Sp. "contemplar" Eng. "contemplate"	Lago Sp. "lago" Eng. "lake"

TABLE A-4.2. Impersonal expressions, proper nouns, verbs and inanimate nouns in the testing set of Experiments 1 and 2. Spanish (Sp.) and English (Eng.) translations are provided below each word.

5. Lexical characteristics of the exposure set across conditions in Experiment 1

		Length Prop. nouns	Length Verbs	Length In. nouns	LD Prop. Nouns	LD Verbs	LD In. nouns	Frequency Prop. nouns	Frequency Verbs	Frequency In. nouns
List 1	*Pl.SS*	5.28	7.04	6.52	0.24	0.52	1.16	15.72	33.25	30.46
		(0.79)	(1.06)	(1.66)	(0.52)	(0.65)	(0.80)	(10.88)	(41.29)	(38.16)
	Pl.DS	5.20	7.20	6.28	0.28	0.56	1.28	13.49	27.62	31.91
		(0.82)	(1.08)	(1.62)	(0.61)	(0.58)	(0.79)	(9.70)	(38.34)	(38.73)
	t-test	*t*(48) = 0.35, *p* = .73	*t*(48) = -0.52, *p* = .60	*t*(48) = 0.52, *p* = .61	*t*(48) = -0.25, *p* = .81	*t*(48) = -0.23, *p* = .82	*t*(48) = -0.53, *p* = .60	*t*(48) = 0.45, *p* = .77	*t*(48) = 0.50, *p* = .62	*t*(48) = -0.13, *p* = .89
List 2	*Pl.SS*	5.20	7.04	6.36	0.20	0.44	1.36	15.78	28.47	29.57
		(0.87)	(1.06)	(1.60)	(0.50)	(0.58)	(0.81)	(10.92)	(37.98)	(37.42)
	Pl.DS	5.28	7.16	6.52	0.24	0.52	1.16	15.72	33.25	30.46
		(0.79)	(1.11)	(1.66)	(0.52)	(0.65)	(0.80)	(10.88)	(41.29)	(38.16)
	t-test	*t*(48) = -0.34, *p* = .73	*t*(48) = 0.39, *p* = .70	*t*(48) = -0.35, *p* = .73	*t*(48) = -0.28, *p* = .78	*t*(48) = -0.46, *p* = .65	*t*(48) = 0.88, *p* = .38	*t*(48) = 0.02, *p* = .99	*t*(48) = -0.43, *p* = .67	*t*(48) = -0.08, *p* = .93
List 3	*Pl.SS*	5.32	7.20	6.64	0.28	0.48	1.40	12.60	36.56	37.98
		(0.90)	(1.26)	(1.73)	(0.54)	(0.59)	(0.76)	(10.28)	(46.85)	(45.39)
	Pl.DS	5.20	7.16	6.36	0.20	0.44	1.36	15.78	28.47	29.57
		(0.87)	(1.11)	(1.60)	(0.50)	(0.58)	(0.81)	(10.92)	(37.98)	(37.42)
	t-test	*t*(48) = 0.48, *p* = .63	*t*(48) = 0.12, *p* = .91	*t*(48) = 0.59, *p* = .55	*t*(48) = 0.54, *p* = .59	*t*(48) = 0.24, *p* = .81	*t*(48) = 0.18, *p* = .86	*t*(48) = -1.06, *p* = .29	*t*(48) = 0.67, *p* = .51	*t*(48) = 0.71, *p* = .48
List 4	*Pl.SS*	5.32	7.20	6.28	0.28	0.56	1.28	13.49	27.62	31.91
		(0.90)	(1.08)	(1.62)	(0.61)	(0.58)	(0.79)	(9.70)	(38.34)	(38.73)
	Pl.DS	5.20	7.20	6.64	0.28	0.48	1.40	12.60	36.56	37.98
		(0.82)	(1.26)	(1.73)	(0.54)	(0.59)	(0.76)	(10.28)	(46.85)	(45.39)
	t-test	*t*(48) = -0.49, *p* = .62	*t*(48) = 0, *p* = 1	*t*(48) = -0.80, *p* = .45	*t*(48) = 0, *p* = 1	*t*(48) = 0.48, *p* = .63	*t*(48) = -0.55, *p* = .59	*t*(48) = 0.32, *p* = .75	*t*(48) = -0.74, *p* = .46	*t*(48) = -0.51, *p* = .61

TABLE A-5.1. Mean length (number of letters), phonological overlap with Spanish translations (Levenshtein distance, LD) and frequency per million in Spanish (SD in brackets) for proper nouns, verbs and inanimate nouns in plausible similar structures (Pl.SS) and plausible dissimilar structures (Pl.DS) in Experiment 1’s four exposure lists. Independent-samples t-tests compare each variable in the two conditions.

		Length Prop. nouns	Length Verbs	Length In. nouns	LD Prop. Nouns	LD Verbs	LD In. nouns	Frequency Prop. nouns	Frequency Verbs	Frequency In. nouns
List 1	*Impl.SS*	5.32 (0.90)	7.20 (1.26)	6.96 (2.11)	0.28 (0.54)	0.48 (0.59)	1.08 (0.86)	12.60 (10.28)	36.56 (46.85)	23.50 (22.43)
	Impl.DS	5.20 (0.87)	7.16 (1.11)	7.04 (2.13)	0.20 (0.50)	0.44 (0.58)	1.24 (0.83)	15.78 (10.92)	28.47 (37.98)	23.80 (22.48)
	t-test	$t(48) = 0.48$, $p = .63$	$t(48) = 0.12$, $p = .91$	$t(48) = -0.13$, $p = .89$	$t(48) = 0.54$, $p = .59$	$t(48) = 0.24$, $p = .81$	$t(48) = -0.67$, $p = .51$	$t(48) = -1.06$, $p = .29$	$t(48) = 0.67$, $p = .51$	$t(48) = -0.05$, $p = .96$
List 2	*Impl.SS*	5.32 (0.90)	7.20 (1.08)	6.56 (2.14)	0.28 (0.61)	0.56 (0.58)	1.08 (0.86)	13.49 (9.70)	27.62 (38.34)	25.31 (23.33)
	Impl.DS	5.20 (0.82)	7.20 (1.26)	6.96 (2.11)	0.28 (0.54)	0.48 (0.59)	1.08 (0.86)	12.60 (10.28)	36.56 (46.85)	23.50 (22.43)
	t-test	$t(48) = -0.49$, $p = .62$	$t(48) = 0$, $p = 1$	$t(48) = -0.66$, $p = .51$	$t(48) = 0$, $p = 1$	$t(48) = 0.48$, $p = .63$	$t(48) = 0$, $p = 1$	$t(48) = 0.32$, $p = .75$	$t(48) = -0.74$, $p = .46$	$t(48) = 0.28$, $p = .78$
List 3	*Impl.SS*	5.28 (0.79)	7.04 (1.06)	6.24 (2.28)	0.24 (0.52)	0.52 (0.65)	1.00 (0.82)	15.72 (10.88)	33.25 (41.29)	28.52 (26.86)
	Impl.DS	5.20 (0.82)	7.20 (1.08)	6.56 (2.14)	0.28 (0.61)	0.56 (0.58)	1.08 (0.86)	13.49 (9.70)	27.62 (38.34)	25.31 (23.33)
	t-test	$t(48) = 0.35$, $p = .73$	$t(48) = -0.52$, $p = .60$	$t(48) = -0.51$, $p = .61$	$t(48) = -0.25$, $p = .81$	$t(48) = -0.23$, $p = .82$	$t(48) = -0.34$, $p = .74$	$t(48) = 0.45$, $p = .77$	$t(48) = 0.50$, $p = .62$	$t(48) = 0.45$, $p = .65$
List 4	*Impl.SS*	5.2 (0.87)	7.04 (1.06)	7.04 (2.13)	0.20 (0.50)	0.44 (0.58)	1.24 (0.83)	15.78 (10.92)	28.47 (37.98)	23.80 (22.48)
	Impl.DS	5.28 (0.79)	7.16 (1.11)	6.24 (2.28)	0.24 (0.52)	0.52 (0.65)	1.00 (0.82)	15.72 (10.88)	33.25 (41.29)	28.52 (26.86)
	t-test	$t(48) = -0.34$, $p = .73$	$t(48) = 0.39$, $p = .70$	$t(48) = 1.28$, $p = .21$	$t(48) = -0.28$, $p = .78$	$t(48) = -0.46$, $p = .65$	$t(48) = 1.03$, $p = .31$	$t(48) = 0.02$, $p = .99$	$t(48) = -0.43$, $p = .67$	$t(48) = 0.67$, $p = .50$

TABLE A-5.2. Mean length (number of letters), phonological overlap with Spanish translations (Levenshtein distance, LD) and frequency per million in Spanish (SD in brackets) for proper nouns, verbs and inanimate nouns in implausible similar structures (Impl.SS) and implausible dissimilar structures (Impl.DS) in Experiment 1’s four exposure lists. Independent-samples t-tests compare each variable in the two conditions.

		Length Prop. nouns	Length Verbs	Length In. nouns	LD Prop. Nouns	LD Verbs	LD In. nouns	Frequency Prop. nouns	Frequency Verbs	Frequency In. nouns
List 1	*Pl.SS*	5.28 (0.79)	7.04 (1.06)	6.52 (1.66)	0.24 (0.52)	0.52 (0.65)	1.16 (0.80)	15.72 (10.88)	33.25 (41.29)	30.46 (38.16)
	Impl.SS	5.32 (0.90)	7.20 (1.26)	6.96 (2.11)	0.28 (0.54)	0.48 (0.59)	1.08 (0.86)	12.60 (10.28)	36.56 (46.85)	23.50 (22.43)
	t-test	$t(48) = -0.17$, $p = .87$	$t(48) = -0.49$, $p = .63$	$t(48) =-0.82$, $p = .42$	$t(48) = -0.27$, $p = .79$	$t(48) = 0.23$, $p = .82$	$t(48) = 0.34$, $p = .74$	$t(48) = 1.05$, $p = .30$	$t(48) = -0.26$, $p = .79$	$t(48) = 0.79$, $p = .44$
List 2	*Pl.SS*	5.2 (0.87)	7.16 (1.11)	6.36 (1.60)	0.20 (0.50)	0.44 (0.58)	1.36 (0.81)	15.78 (10.92)	28.47 (37.98)	29.57 (37.42)
	Impl.SS	5.2 (0.82)	7.20 (1.08)	6.56 (2.14)	0.28 (0.61)	0.56 (0.58)	1.08 (0.86)	13.49 (9.70)	27.62 (38.34)	25.31 (23.33)
	t-test	$t(48) = 0$, $p = 1$	$t(48) = -0.13$, $p = .90$	$t(48) = -1.42$, $p = .16$	$t(48) = -0.51$, $p = .62$	$t(48) = -0.73$, $p = .47$	$t(48) = 0.17$, $p = .86$	$t(48) = 0.78$, $p = .44$	$t(48) = 0.08$, $p = .94$	$t(48) = 0.91$, $p = .37$
List 3	*Pl.SS*	5.32 (0.90)	7.20 (1.26)	6.64 (1.73)	0.28 (0.54)	0.48 (0.59)	1.40 (0.76)	12.60 (10.28)	36.56 (46.85)	37.98 (45.39)
	Impl.SS	5.28 (0.79)	7.04 (1.06)	6.24 (2.28)	0.24 (0.52)	0.52 (0.65)	1.00 (0.82)	15.72 (10.88)	33.25 (41.29)	28.52 (26.86)
	t-test	$t(48) = 0.17$, $p = .87$	$t(48) = 0.49$, $p = .63$	$t(48) = 0.70$, $p = .49$	$t(48) = 0.27$, $p = .79$	$t(48) = -0.23$, $p = .82$	$t(48) = 1.79$, $p = .08$	$t(48) = -1.05$, $p = .30$	$t(48) = 0.26$, $p = .79$	$t(48) = 0.90$, $p = .37$
List 4	*Pl.SS*	5.2 (0.82)	7.20 (1.08)	6.28 (1.62)	0.28 (0.61)	0.56 (0.58)	1.28 (0.79)	13.49 (9.70)	27.62 (38.34)	31.91 (38.73)
	Impl.SS	5.2. (0.87)	7.16 (1.11)	7.04 (2.13)	0.20 (0.50)	0.44 (0.58)	1.24 (0.83)	15.78 (10.92)	28.47 (37.98)	23.80 (22.48)
	t-test	$t(48) = 0$, $p = 1$	$t(48) = 0.13$, $p = .90$	$t(48) = 0.37$, $p = .71$	$t(48) = 0.51$, $p = .62$	$t(48) = 0.73$, $p = .47$	$t(48) = 1.18$, $p = .24$	$t(48) = -0.78$, $p = .44$	$t(48) = -0.08$, $p = .94$	$t(48) = 0.48$, $p = .63$

TABLE A-5.3. Mean length (number of letters), phonological overlap with Spanish translations (Levenshtein distance, LD) and frequency per million in Spanish (SD in brackets) for proper nouns, verbs and inanimate nouns in plausible similar structures (Pl.SS) and implausible similar structures (Impl.SS) in Experiment 1's four exposure lists. Independent-samples t-tests compare each variable in the two conditions.

		Length Prop. nouns	Length Verbs	Length In. nouns	LD Prop. Nouns	LD Verbs	LD In. nouns	Frequency Prop. nouns	Frequency Verbs	Frequency In. nouns
List 1	*Pl.DS*	5.2	7.20	6.28	0.28	0.56	1.28	13.49	27.62	31.91
		(0.82)	(1.08)	(1.62)	(0.61)	(0.58)	(0.79)	(9.70)	(38.34)	(38.73)
	Impl.DS	5.2	7.16	7.04	0.20	0.44	1.24	15.78	28.47	23.80
		(0.87)	(1.11)	(2.13)	(0.50)	(0.58)	(0.83)	(10.92)	(37.98)	(22.48)
	t-test	$t(48) = 0$, $p = 1$	$t(48) = 0.13$, $p = .90$	$t(48) = -1.42$, $p = .16$	$t(48) = 0.51$, $p = .62$	$t(48) = 0.73$, $p = .47$	$t(48) = -0.17$, $p = .86$	$t(48) = -0.78$, $p = .44$	$t(48) = -0.08$, $p = .94$	$t(48) = 0.91$, $p = .37$
List 2	*Pl.DS*	5.28	7.04	6.64	0.24	0.52	1.40	15.72	33.25	37.98
		(0.79)	(1.06)	(1.73)	(0.52)	(0.65)	(0.76)	(10.88)	(41.29)	(45.39)
	Impl.DS	5.32	7.20	6.24	0.28	0.48	1.00	12.60	36.56	28.52
		(0.90)	(1.26)	(2.28)	(0.54)	(0.59)	(0.82)	(10.28)	(46.85)	(26.86)
	t-test	$t(48) = -0.17$, $p = .87$	$t(48) = -0.49$, $p = .63$	$t(48) = 0.70$, $p = .49$	$t(48) = -0.27$, $p = .79$	$t(48) = 0.23$, $p = .82$	$t(48) = 1.79$, $p = .08$	$t(48) = 1.05$, $p = .30$	$t(48) = -0.26$, $p = .79$	$t(48) = 0.90$, $p = .37$
List 3	*Pl.DS*	5.2	7.16	6.36	0.20	0.44	1.36	15.78	28.47	29.57
		(0.87)	(1.11)	(1.60)	(0.50)	(0.58)	(0.81)	(10.92)	(37.98)	(37.42)
	Impl.DS	5.2	7.20	6.56	0.28	0.56	1.08	13.49	27.62	25.31
		(0.82)	(1.08)	(2.14)	(0.61)	(0.58)	(0.86)	(9.70)	(38.34)	(23.33)
	t-test	$t(48) = 0$, $p = 1$	$t(48) = -0.13$, $p = .90$	$t(48) = -0.37$, $p = .71$	$t(48) = -0.51$, $p = .62$	$t(48) = -0.73$, $p = .47$	$t(48) = 1.18$, $p = .24$	$t(48) = 0.78$, $p = .44$	$t(48) = 0.08$, $p = .94$	$t(48) = 0.48$, $p = .63$
List 4	*Pl.DS*	5.32	7.20	6.52	0.28	0.48	1.16	12.60	36.56	30.46
		(0.90)	(1.26)	(1.66)	(0.54)	(0.59)	(0.80)	(10.28)	(46.85)	(38.16)
	Impl.DS	5.28	7.04	6.96	0.24	0.52	1.08	15.72	33.25	23.50
		(0.79)	(1.06)	(2.11)	(0.52)	(0.65)	(0.86)	(10.88)	(41.29)	(22.43)
	t-test	$t(48) = 0.17$, $p = .87$	$t(48) = 0.49$, $p = .63$	$t(48) = -0.82$, $p = .42$	$t(48) = 0.27$, $p = .79$	$t(48) = -0.23$, $p = .82$	$t(48) = 0.34$, $p = .74$	$t(48) = -1.05$, $p = .30$	$t(48) = 0.26$, $p = .79$	$t(48) = 0.79$, $p = .44$

TABLE A-5.4. Mean length (number of letters), phonological overlap with Spanish translations (Levenshtein distance, LD) and frequency per million in Spanish (SD in brackets) for proper nouns, verbs and inanimate nouns in plausible dissimilar structures (Pl.DS) and implausible dissimilar structures (Impl.DS) in Experiment 1's four exposure lists. Independent-samples t-tests compare each variable in the two conditions.

6. Experimental materials used in Experiment 1

Exposure set

Galician-based sentences constituting the exposure set in Experiment 1. Sentences are presented in groups of four, corresponding to the four conditions in the exposure phase: a. Plausible similar structure, b. Plausible dissimilar structure, c. Implausible similar structure and d. Implausible dissimilar structure. English translations are provided for each group of sentences. The translation is shared for plausible sentences (a and b) and implausible sentences (c and d).

(1)
a. É importante que Pedro repare a radio.
b. É importante Pedro reparar a radio.
c. É importante que Pedro repare o tornado.
d. É importante Pedro reparar o tornado.
"It is important that Pedro fixes the radio / the tornado."

(2)
a. É posible que Mónica apague a consola.
b. É posible Mónica apagar a consola.
c. É posible que Mónica apague o vento.
d. É posible Mónica apagar o vento.
"It is possible that Mónica turns off the console / the wind."

(3)
a. É probable que Ángel acenda o móbil.
b. É probable Ángel acender o móbil.
c. É probable que Ángel acenda o sol.
d. É probable Ángel acender o sol.
"It is probable that Ángel turns on the mobile / the sun."

(4)
a. É sorprendente que Isabel venda o portátil.
b. É sorprendente Isabel vender o portátil.
c. É sorprendente que Isabel venda o mar.
d. É sorprendente Isabel vender o mar.
"It is surprising that Isabel sells the laptop / the sea."

(5)
a. É necesario que Pablo compre o televisor.
b. É necesario Pablo comprar o televisor.
c. É necesario que Pablo compre a nube.
d. É necesario Pablo comprar a nube.
"It is necessary that Pablo buys the television / the cloud."

(6) a. É importante que Lucía saboree a torta.
b. É importante Lucía saborear a torta.
c. É importante que Lucía saboree o río.
d. É importante Lucía saborear o río.
"It is important that Lucía savors the cake / the river."

(7) a. É posible que Luis corte a carne.
b. É posible Luis cortar a carne.
c. É posible que Luis corte o monte.
d. É posible Luis cortar o monte.
"It is possible that Luis cuts the meat / the mount."

(8) a. É probable que Paula cociñe o pan.
b. É probable Paula cociñar o pan.
c. É probable que Paula cociñe a montaña.
d. É probable Paula cociñar a montaña.
"It is probable that Paula bakes the bread / the mountain."

(9) a. É sorprendente que Jorge mastigue o brócoli.
b. É sorprendente Jorge mastigar o brócoli.
c. É sorprendente que Jorge mastigue o estanque.
d. É sorprendente Jorge mastigar o estanque.
"It is surprising that Jorge chews the broccoli / the pond."

(10) a. É necesario que Alicia conxele a cenoria.
b. É necesario Alicia conxelar a cenoria.
c. É necesario que Alicia conxele a galaxia.
d. É necesario Alicia conxelar a galaxia.
"It is necessary that Alicia freezes the carrot / the galaxy."

(11) a. É importante que Alberto gañe o concurso.
b. É importante Alberto gañar o concurso.
c. É importante que Alberto gañe o universo.
d. É importante Alberto gañar o universo.
"It is important that Alberto wins the contest / the universe."

(12) a. É posible que Silvia perda a proba.
b. É posible Silvia perder a proba.
c. É posible que Silvia perda o océano.
d. É posible Silvia perder o océano.
"It is possible that Silvia loses the test / the ocean."

(13) a. É probable que Adrián financie o torneo.
b. É probable Adrián financiar o torneo.
c. É probable que Adrián financie o amañecer.
d. É probable Adrián financiar o amañecer.
"It is probable that Adrián finances the tournament / the sunrise."

(14) a. É sorprendente que Irene organice a carreira.
b. É sorprendente Irene organizar a carreira.
c. É sorprendente que Irene organice o barranco.
d. É sorprendente Irene organizar o barranco.
"It is surprising that Irene organizes the race / the gorge."

(15) a. É necesario que Iván cancele o campionato.
b. É necesario Iván cancelar o campionato.
c. É necesario que Iván cancele o precipicio.
d. É necesario Iván cancelar o precipicio.
"It is necessary that Iván cancels the championship / the cliff."

(16) a. É importante que Rosa devolva a xoia.
b. É importante Rosa devolver a xoia.
c. É importante que Rosa devolva a tormenta.
d. É importante Rosa devolver a tormenta.
"It is important that Rosa returns the jewel / the storm."

(17) a. É posible que Raúl garde a pulseira.
b. É posible Raúl gardar a pulseira.
c. É posible que Raúl garde o furacán.
d. É posible Raúl gardar o furacán.
"It is possible that Raúl stores the bracelet / the hurricane."

(18) a. É probable que Andrea encargue o abrigo.
b. É probable Andrea encargar o abrigo.
c. É probable que Andrea encargue o terremoto.
d. É probable Andrea encargar o terremoto.
"It is probable that Andrea orders the coat / the earthquake."

(19) a. É sorprendente que Óscar envolva o reloxo.
b. É sorprendente Óscar envolver o reloxo.
c. É sorprendente que Óscar envolva o incendio.
d. É sorprendente Óscar envolver o incendio.
"It is surprising that Óscar wraps the watch / the fire."

(20) a. É necesario que Carmen envíe o xersei.
b. É necesario Carmen enviar o xersei.
c. É necesario que Carmen envíe a inundación.
d. É necesario Carmen enviar a inundación.
"It is necessary that Carmen sends the sweater / the flood."

(21) a. É importante que Pablo venda o móbil.
b. É importante Pablo vender o móbil.
c. É importante que Pablo venda o sol.
d. É importante Pablo vender o sol.
"It is important that Pablo sells the mobile / the sun."

(22) a. É posible que Pedro compre o portátil.
b. É posible Pedro comprar o portátil.
c. É posible que Pedro compre o mar.
d. É posible Pedro comprar o mar.
"It is possible that Pedro buys the laptop / the sea."

(23) a. É probable que Mónica repare o televisor.
b. É probable Mónica reparar o televisor.
c. É probable que Mónica repare a nube.
d. É probable Mónica reparar a nube.
"It is probable that Mónica fixes the television / the cloud."

(24) a. É sorprendente que Ángel apague a radio.
b. É sorprendente Ángel apagar a radio.
c. É sorprendente que Ángel apague o tornado.
d. É sorprendente Ángel apagar o tornado.
"It is surprising that Ángel turns off the radio / the tornado."

(25) a. É necesario que Isabel acenda a consola.
b. É necesario Isabel acender a consola.
c. É necesario que Isabel acenda o vento.
d. É necesario Isabel acender o vento.
"It is necessary that Isabel turns on the console / the wind."

(26) a. É importante que Alicia mastigue o pan.
b. É importante Alicia mastigar o pan.
c. É importante que Alicia mastigue a montaña.
d. É importante Alicia mastigar a montaña.
"It is important that Alicia chews the bread / the mountain."

(27) a. É posible que Lucía conxele o brócoli.
b. É posible Lucía conxelar o brócoli.
c. É posible que Lucía conxele o estanque.
d. É posible Lucía conxelar o estanque.
"It is possible that Lucía freezes the broccoli / the pond."

(28) a. É probable que Luis saboree a cenoria.
b. É probable Luis saborear a cenoria.
c. É probable que Luis saboree a galaxia.
d. É probable Luis saborear a galaxia.
"It is probable that Luis savors the carrot / the galaxy."

(29) a. É sorprendente que Paula corte a torta.
b. É sorprendente Paula cortar a torta.
c. É sorprendente que Paula corte o río.
d. É sorprendente Paula cortar o río.
"It is surprising that Paula cuts the cake / the river."

(30) a. É necesario que Jorge cociñe a carne.
b. É necesario Jorge cociñar a carne.
c. É necesario que Jorge cociñe o monte.
d. É necesario Jorge cociñar o monte.
"It is necessary that Jorge cooks the meat / the mount."

(31) a. É importante que Iván organice o torneo.
b. É importante Iván organizar o torneo.
c. É importante que Iván organice o amañecer.
d. É importante Iván organizar o amañecer.
"It is important that Iván organizes the tournament / the sunrise."

(32) a. É posible que Alberto cancele a carreira.
b. É posible Alberto cancelar a carreira.
c. É posible que Alberto cancele o barranco.
d. É posible Alberto cancelar o barranco.
"It is possible that Alberto cancels the race / the gorge."

(33) a. É probable que Silvia gañe o campionato.
b. É probable Silvia gañar o campionato.
c. É probable que Silvia gañe o precipicio.
d. É probable Silvia gañar o precipicio.
"It is probable that Silvia wins the championship / the cliff."

(34) a. É sorprendente que Adrián perda o concurso.
b. É sorprendente Adrián perder o concurso.
c. É sorprendente que Adrián perda o universo.
d. É sorprendente Adrián perder o universo.
"It is surprising that Adrián loses the contest / the universe."

(35) a. É necesario que Irene financie a proba.
b. É necesario Irene financiar a proba.
c. É necesario que Irene financie o océano.
d. É necesario Irene financiar o océano.
"It is necessary that Irene finances the test / the ocean."

(36) a. É importante que Carmen envolva o abrigo.
b. É importante Carmen envolver o abrigo.
c. É importante que Carmen envolva o terremoto.
d. É importante Carmen envolver o terremoto.
"It is important that Carmen wraps the coat / the earthquake."

(37) a. É posible que Rosa envíe o reloxo.
b. É posible Rosa enviar o reloxo.
c. É posible que Rosa envíe o incendio.
d. É posible Rosa enviar o incendio.
"It is possible that Rosa sends the watch / the fire."

(38) a. É probable que Raúl devolva o xersei.
b. É probable Raúl devolver o xersei.
c. É probable que Raúl devolva a inundación.
d. É probable Raúl devolver a inundación.
"It is probable that Raúl returns the sweater / the flood."

(39) a. É sorprendente que Andrea garde a xoia.
b. É sorprendente Andrea gardar a xoia.
c. É sorprendente que Andrea garde a tormenta.
d. É sorprendente Andrea gardar a tormenta.
"It is surprising that Andrea stores the jewel / the storm."

(40) a. É necesario que Óscar encargue a pulseira.
b. É necesario Óscar encargar a pulseira.
c. É necesario que Óscar encargue o furacán.
d. É necesario Óscar encargar o furacán.
"It is necessary that Óscar orders the bracelet / the hurricane."

(41) a. É importante que Isabel apague o televisor.
b. É importante Isabel apagar o televisor.
c. É importante que Isabel apague a nube.
d. É importante Isabel apagar a nube.
"It is important that Isabel turns off the television / the cloud."

(42) a. É posible que Pablo acenda a radio.
b. É posible Pablo acender a radio.
c. É posible que Pablo acenda o tornado.
d. É posible Pablo acender o tornado.
"It is possible that Pablo turns on the radio / the tornado."

(43) a. É probable que Pedro venda a consola.
b. É probable Pedro vender a consola.
c. É probable que Pedro venda o vento.
d. É probable Pedro vender o vento.
"It is probable that Pedro sells the console / the wind."

(44) a. É sorprendente que Mónica compre o móbil.
b. É sorprendente Mónica comprar o móbil.
c. É sorprendente que Mónica compre o sol.
d. É sorprendente Mónica comprar o sol.
"It is surprising that Mónica buys the mobile / the sun."

(45) a. É necesario que Ángel repare o portátil.
b. É necesario Ángel reparar o portátil.
c. É necesario que Ángel repare o mar.
d. É necesario Ángel reparar o mar.
"It is necessary that Ángel fixes the laptop / the sea."

(46) a. É importante que Jorge corte a cenoria.
b. É importante Jorge cortar a cenoria.
c. É importante que Jorge corte a galaxia.
d. É importante Jorge cortar a galaxia.
"It is important that Jorge cuts the carrot / the galaxy."

(47) a. É posible que Alicia cociñe a torta.
b. É posible Alicia cociñar a torta.
c. É posible que Alicia cociñe o río.
d. É posible Alicia cociñar o río.
"It is possible that Alicia bakes the cake / the river."

(48) a. É probable que Lucía mastigue a carne.
b. É probable Lucía mastigar a carne.
c. É probable que Lucía mastigue o monte.
d. É probable Lucía mastigar o monte.
"It is probable that Lucía chews the meat / the mount."

(49) a. É sorprendente que Luis conxele o pan.
b. É sorprendente Luis conxelar o pan.
c. É sorprendente que Luis conxele a montaña.
d. É sorprendente Luis conxelar a montaña.
"It is surprising that Luis freezes the bread / the mountain."

(50) a. É necesario que Paula saboree o brócoli.
b. É necesario Paula saborear o brócoli.
c. É necesario que Paula saboree o estanque.
d. É necesario Paula saborear o estanque.
"It is necessary that Paula savors the broccoli / the pond."

(51) a. É importante que Irene perda o campionato.
b. É importante Irene perder o campionato.
c. É importante que Irene perda o precipicio.
d. É importante Irene perder o precipicio.
"It is important that Irene loses the championship / the cliff."

(52) a. É posible que Iván financie o concurso.
b. É posible Iván financiar o concurso.
c. É posible que Iván financie o universo.
d. É posible Iván financiar o universo.
"It is possible that Iván finances the contest / the universe."

(53) a. É probable que Alberto organice a proba.
b. É probable Alberto organizar a proba.
c. É probable que Alberto organice o océano.
d. É probable Alberto organizar o océano.
"It is probable that Alberto organizes the test / the ocean."

(54) a. É sorprendente que Silvia cancele o torneo.
b. É sorprendente Silvia cancelar o torneo.
c. É sorprendente que Silvia cancele o amañecer.
d. É sorprendente Silvia cancelar o amañecer.
"It is surprising that Silvia cancels the tournament / the sunrise."

(55) a. É necesario que Adrián gañe a carreira.
b. É necesario Adrián gañar a carreira.
c. É necesario que Adrián gañe o barranco.
d. É necesario Adrián gañar o barranco.
"It is necessary that Adrián wins the race / the gorge."

(56) a. É importante que Óscar garde o xersei.
b. É importante Óscar gardar o xersei.
c. É importante que Óscar garde a inundación.
d. É importante Óscar gardar a inundación.
"It is important that Óscar stores the sweater / the flood."

(57) a. É posible que Carmen encargue a xoia.
b. É posible Carmen encargar a xoia.
c. É posible que Carmen encargue a tormenta.
d. É posible Carmen encargar a tormenta.
"It is possible that Carmen orders the jewel / the storm."

(58) a. É probable que Rosa envolva a pulseira.
b. É probable Rosa envolver a pulseira.
c. É probable que Rosa envolva o furacán.
d. É probable Rosa envolver o furacán.
"It is probable that Rosa wraps the bracelet / the hurricane."

(59) a. É sorprendente que Raúl envíe o abrigo.
b. É sorprendente Raúl enviar o abrigo.
c. É sorprendente que Raúl envíe o terremoto.
d. É sorprendente Raúl enviar o terremoto.
"It is surprising that Raúl sends the coat / the earthquake."

(60) a. É necesario que Andrea devolva o reloxo.
b. É necesario Andrea devolver o reloxo.
c. É necesario que Andrea devolva o incendio.
d. É necesario Andrea devolver o incendio.
"It is necessary that Andrea returns the watch / the fire."

(61) a. É importante que Ángel compre a consola.
b. É importante Ángel comprar a consola.
c. É importante que Ángel compre o vento.
d. É importante Ángel comprar o vento.
"It is important that Ángel buys the console / the wind."

(62) a. É posible que Isabel repare o móbil.
b. É posible Isabel reparar o móbil.
c. É posible que Isabel repare o sol.
d. É posible Isabel reparar o sol.
"It is possible that Isabel fixes the mobile / the sun."

(63) a. É probable que Pablo apague o portátil.
b. É probable Pablo apagar o portátil.
c. É probable que Pablo apague o mar.
d. É probable Pablo apagar o mar.
"It is probable that Pablo turns off the laptop / the sea."

(64) a. É sorprendente que Pedro acenda o televisor.
b. É sorprendente Pedro acender o televisor.
c. É sorprendente que Pedro acenda a nube.
d. É sorprendente Pedro acender a nube.
"It is surprising that Pedro turns on the television / the cloud."

(65) a. É necesario que Mónica venda a radio.
b. É necesario Mónica vender a radio.
c. É necesario que Mónica venda o tornado.
d. É necesario Mónica vender o tornado.
"It is necessary that Mónica sells the radio / the tornado."

(66) a. É importante que Paula conxele a carne.
b. É importante Paula conxelar a carne.
c. É importante que Paula conxele o monte.
d. É importante Paula conxelar o monte.
"It is important that Paula freezes the meat / the mount."

(67) a. É posible que Jorge saboree o pan.
b. É posible Jorge saborear o pan.
c. É posible que Jorge saboree a montaña.
d. É posible Jorge saborear a montaña.
"It is possible that Jorge savors the bread / the mountain."

(68) a. É probable que Alicia corte o brócoli.
b. É probable Alicia cortar o brócoli.
c. É probable que Alicia corte o estanque.
d. É probable Alicia cortar o estanque.
"It is probable that Alicia cuts the broccoli / the pond."

(69) a. É sorprendente que Lucía cociñe a cenoria.
b. É sorprendente Lucía cociñar a cenoria.
c. É sorprendente que Lucía cociñe a galaxia.
d. É sorprendente Lucía cociñar a galaxia.
"It is surprising that Lucía cooks the carrot / the galaxy."

(70) a. É necesario que Luis mastigue a torta.
b. É necesario Luis mastigar a torta.
c. É necesario que Luis mastigue o río.
d. É necesario Luis mastigar o río.
"It is necessary that Luis chews the cake / the river."

(71) a. É importante que Adrián cancele a proba.
b. É importante Adrián cancelar a proba.
c. É importante que Adrián cancele o océano.
d. É importante Adrián cancelar o océano.
"It is important that Adrián cancels the test / the ocean."

(72) a. É posible que Irene gañe o torneo.
b. É posible Irene gañar o torneo.
c. É posible que Irene gañe o amañecer.
d. É posible Irene gañar o amañecer.
"It is possible that Irene wins the tournament / the sunrise."

(73) a. É probable que Iván perda a carreira.
b. É probable Iván perder a carreira.
c. É probable que Iván perda o barranco.
d. É probable Iván perder o barranco.
"It is probable that Iván loses the race / the gorge."

(74) a. É sorprendente que Alberto financie o campionato.
b. É sorprendente Alberto financiar o campionato.
c. É sorprendente que Alberto financie o precipicio.
d. É sorprendente Alberto financiar o precipicio.
"It is surprising that Alberto finances the championship / the cliff."

(75) a. É necesario que Silvia organice o concurso.
b. É necesario Silvia organizar o concurso.
c. É necesario que Silvia organice o universo.
d. É necesario Silvia organizar o universo.
"It is necessary that Silvia organizes the contest / the universe."

(76) a. É importante que Andrea envíe a pulseira.
b. É importante Andrea enviar a pulseira.
c. É importante que Andrea envíe o furacán.
d. É importante Andrea enviar o furacán.
"It is important that Andrea sends the bracelet / the hurricane."

(77) a. É posible que Óscar devolva o abrigo.
b. É posible Óscar devolver o abrigo.
c. É posible que Óscar devolva o terremoto.
d. É posible Óscar devolver o terremoto.
"It is possible that Óscar returns the coat / the earthquake."

(78) a. É probable que Carmen garde o reloxo.
b. É probable Carmen gardar o reloxo.
c. É probable que Carmen garde o incendio.
d. É probable Carmen gardar o incendio.
"It is probable that Carmen stores the watch / the fire."

(79) a. É sorprendente que Rosa encargue o xersei.
b. É sorprendente Rosa encargar o xersei.
c. É sorprendente que Rosa encargue a inundación.
d. É sorprendente Rosa encargar a inundación.
"It is surprising that Rosa orders the sweater / the flood."

(80) a. É necesario que Raúl envolva a xoia.
b. É necesario Raúl envolver a xoia.
c. É necesario que Raúl envolva a tormenta.
d. É necesario Raúl envolver a tormenta.
"It is necessary that Raúl wraps the jewel / the storm."

(81) a. É importante que Mónica acenda o portátil.
b. É importante Mónica acender o portátil.
c. É importante que Mónica acenda o mar.
d. É importante Mónica acender o mar.
"It is important that Mónica turns on the laptop / the sea."

(82) a. É posible que Ángel venda o televisor.
b. É posible Ángel vender o televisor.
c. É posible que Ángel venda a nube.
d. É posible Ángel vender a nube.
"It is possible that Ángel sells the television / the cloud."

(83) a. É probable que Isabel compre a radio.
b. É probable Isabel comprar a radio.
c. É probable que Isabel compre o tornado.
d. É probable Isabel comprar o tornado.
"It is probable that Isabel buys the radio / the tornado."

(84) a. É sorprendente que Pablo repare a consola.
b. É sorprendente Pablo reparar a consola.
c. É sorprendente que Pablo repare o vento.
d. É sorprendente Pablo reparar o vento.
"It is surprising that Pablo fixes the console / the wind."

(85) a. É necesario que Pedro apague o móbil.
b. É necesario Pedro apagar o móbil.
c. É necesario que Pedro apague o sol.
d. É necesario Pedro apagar o sol.
"It is necessary that Pedro turns off the mobile / the sun."

(86) a. É importante que Luis cociñe o brócoli.
b. É importante Luis cociñar o brócoli.
c. É importante que Luis cociñe o estanque.
d. É importante Luis cociñar o estanque.
"It is important that Luis cooks the broccoli / the pond."

(87) a. É posible que Paula mastigue a cenoria.
b. É posible Paula mastigar a cenoria.
c. É posible que Paula mastigue a galaxia.
d. É posible Paula mastigar a galaxia.
"It is possible that Paula chews the carrot / the galaxy."

(88) a. É probable que Jorge conxele a torta.
b. É probable Jorge conxelar a torta.
c. É probable que Jorge conxele o río.
d. É probable Jorge conxelar o río.
"It is probable that Jorge freezes the cake / the river."

(89) a. É sorprendente que Alicia saboree a carne.
b. É sorprendente Alicia saborear a carne.
c. É sorprendente que Alicia saboree o monte.
d. É sorprendente Alicia saborear o monte.
"It is surprising that Alicia savors the meat / the mount."

(90) a. É necesario que Lucía corte o pan.
b. É necesario Lucía cortar o pan.
c. É necesario que Lucía corte a montaña.
d. É necesario Lucía cortar a montaña.
"It is necessary that Lucía cuts the bread / the mountain."

(91) a. É importante que Silvia financie a carreira.
b. É importante Silvia financiar a carreira.
c. É importante que Silvia financie o barranco.
d. É importante Silvia financiar o barranco.
"It is important that Silvia finances the race / the gorge."

(92) a. É posible que Adrián organice o campionato.
b. É posible Adrián organizar o campionato.
c. É posible que Adrián organice o precipicio.
d. É posible Adrián organizar o precipicio.
"It is possible that Adrián organizes the championship / the cliff."

(93) a. É probable que Irene cancele o concurso.
b. É probable Irene cancelar o concurso.
c. É probable que Irene cancele o universo.
d. É probable Irene cancelar o universo.
"It is probable that Irene cancels the contest / the universe."

(94) a. É sorprendente que Iván gañe a proba.
b. É sorprendente Iván gañar a proba.
c. É sorprendente que Iván gañe o océano.
d. É sorprendente Iván gañar o océano.
"It is surprising that Iván wins the test / the ocean."

(95) a. É necesario que Alberto perda o torneo.
b. É necesario Alberto perder o torneo.
c. É necesario que Alberto perda o amañecer.
d. É necesario Alberto perder o amañecer.
"It is necessary that Alberto loses the tournament / the sunrise."

(96) a. É importante que Raúl encargue o reloxo.
b. É importante Raúl encargar o reloxo.
c. É importante que Raúl encargue o incendio.
d. É importante Raúl encargar o incendio.
"It is important that Raúl orders the watch / the fire."

(97) a. É posible que Andrea envolva o xersei.
b. É posible Andrea envolver o xersei.
c. É posible que Andrea envolva a inundación.
d. É posible Andrea envolver a inundación.
"It is possible that Andrea wraps the sweater / the flood."

(98) a. É probable que Óscar envíe a xoia.
b. É probable Óscar enviar a xoia.
c. É probable que Óscar envíe a tormenta.
d. É probable Óscar enviar a tormenta.
"It is probable that Óscar sends the jewel / the storm."

(99) a. É sorprendente que Carmen devolva a pulseira.
b. É sorprendente Carmen devolver a pulseira.
c. É sorprendente que Carmen devolva o furacán.
d. É sorprendente Carmen devolver o furacán.
"It is surprising that Carmen returns the bracelet / the hurricane."

(100) a. É necesario que Rosa garde o abrigo.
b. É necesario Rosa gardar o abrigo.
c. É necesario que Rosa garde o terremoto.
d. É necesario Rosa gardar o terremoto.
"It is necessary that Rosa stores the coat / the earthquake."

<u>Testing set</u>

Galician-based sentences constituting the testing set in Experiment 1. Sentences are presented in groups of four, corresponding to the four conditions in the testing phase: a. Grammatical similar structure, b. Grammatical dissimilar structure, c. Ungrammatical similar structure and d. Ungrammatical dissimilar structure. I provide the English translation shared for each group of sentences (intended meaning for all sentences irrespective of their grammaticality).

(1) a. É importante que Antonio firme a carta.
b. É importante Antonio firmar a carta.
c. *É importante que Antonio firmar a carta.
d. *É importante Antonio firme a carta.
"It is important that Antonio signs the letter."

(2) a. É posible que Sofía modifique o informe.
b. É posible Sofía modificar o informe.
c. *É posible que Sofía modificar o informe.
d. *É posible Sofía modifique o informe.
"It is possible that Sofía modifies the report."

(3) a. É probable que José reciba a postal.
b. É probable José recibir a postal.
c. *É probable que José recibir a postal.
d. *É probable José reciba a postal.
"It is probable that José receives the postcard."

(4) a. É sorprendente que Ana corrixa o correo.
b. É sorprendente Ana corrixir o correo.
c. *É sorprendente que Ana corrixir o correo.
d. *É sorprendente Ana corrixa o correo.
"It is surprising that Ana corrects the email."

(5) a. É necesario que Manuel escriba a novela.
b. É necesario Manuel escribir a novela.
c. *É necesario que Manuel escribir a novela.
d. *É necesario Manuel escriba a novela.
"It is necessary that Manuel writes the novel."

(6) a. É importante que Laura consulte a factura.
b. É importante Laura consultar a factura.
c. *É importante que Laura consultar a factura.
d. *É importante Laura consulte a factura.
"It is important that Laura checks the invoice."

(7) a. É posible que Francisco lea o libro.
b. É posible Francisco ler o libro.
c. *É posible que Francisco ler o libro.
d. *É posible Francisco lea o libro.
"It is possible that Francisco reads the book."

(8) a. É probable que Victoria redacte o comunicado.
b. É probable Victoria redactar o comunicado.
c. *É probable que Victoria redactar o comunicado.
d. *É probable Victoria redacte o comunicado.
"It is probable that Victoria writes the statement."

(9) a. É sorprendente que Juan revise a proposta.
b. É sorprendente Juan revisar a proposta.
c. *É sorprendente que Juan revisar a proposta.
d. *É sorprendente Juan revise a proposta.
"It is surprising that Juan revises the proposal."

(10) a. É necesario que Marta acepte a lei.
b. É necesario Marta aceptar a lei.
c. *É necesario que Marta aceptar a lei.
d. *É necesario Marta acepte a lei.
"It is necessary that Marta accepts the law."

(11) a. É importante que Alejandro pinte a casa.
b. É importante Alejandro pintar a casa.
c. *É importante que Alejandro pintar a casa.
d. *É importante Alejandro pinte a casa.
"It is important that Alejandro paints the house."

(12) a. É posible que Elena limpe a habitación.
b. É posible Elena limpar a habitación.
c. *É posible que Elena limpar a habitación.
d. *É posible Elena limpe a habitación.
"It is possible that Elena cleans the room."

(13) a. É probable que Javier ordene a cociña.
b. É probable Javier ordenar a cociña.
c. *É probable que Javier ordenar a cociña.
d. *É probable Javier ordene a cociña.
"It is probable that Javier tidies up the kitchen."

(14) a. É sorprendente que Sara contemple o lago.
b. É sorprendente Sara contemplar o lago.
c. *É sorprendente que Sara contemplar o lago.
d. *É sorprendente Sara contemple o lago.
"It is surprising that Sara contemplates the lake."

(15) a. É necesario que Salvador reforme o apartamento.
b. É necesario Salvador reformar o apartamento.
c. *É necesario que Salvador reformar o apartamento.
d. *É necesario Salvador reforme o apartamento.
"It is necessary that Salvador renovates the apartment."

(16) a. É importante que Ángela describa a paisaxe.
b. É importante Ángela describir a paisaxe.
c. *É importante que Ángela describir a paisaxe.
d. *É importante Ángela describa a paisaxe.
"It is important that Ángela describes the landscape."

(17) a. É posible que Carlos observe o cadro.
b. É posible Carlos observar o cadro.
c. *É posible que Carlos observar o cadro.
d. *É posible Carlos observe o cadro.
"It is possible that Carlos observes the painting."

(18) a. É probable que Julia admire a escultura.
b. É probable Julia admirar a escultura.
c. *É probable que Julia admirar a escultura.
d. *É probable Julia admire a escultura.
"It is probable that Julia admires the sculpture."

(19) a. É sorprendente que Miguel fotografe o xardín.
b. É sorprendente Miguel fotografar o xardín.
c. *É sorprendente que Miguel fotografar o xardín.
d. *É sorprendente Miguel fotografe o xardín.
"It is surprising that Miguel photographs the garden."

(20) a. É necesario que Alba recolla o almacén.
b. É necesario Alba recoller o almacén.
c. *É necesario que Alba recoller o almacén.
d. *É necesario Alba recolla o almacén.
"It is necessary that Alba cleans up the warehouse."

(21) a. É importante que Manuel corrixa a postal.
b. É importante Manuel corrixir a postal.
c. *É importante que Manuel corrixir a postal.
d. *É importante Manuel corrixa a postal.
"It is important that Manuel corrects the postcard."

(22) a. É posible que Antonio escriba o correo.
b. É posible Antonio escribir o correo.
c. *É posible que Antonio escribir o correo.
d. *É posible Antonio escriba o correo.
"It is possible that Antonio writes the email."

(23) a. É probable que Sofía firme a novela.
b. É probable Sofía firmar a novela.
c. *É probable que Sofía firmar a novela.
d. *É probable Sofía firme a novela.
"It is probable that Sofía signs the novel."

(24) a. É sorprendente que José modifique a factura.
b. É sorprendente José modificar a factura.
c. *É sorprendente que José modificar a factura.
d. *É sorprendente José modifique a factura.
"It is surprising that José modifies the invoice."

(25) a. É necesario que Ana reciba o libro.
b. É necesario Ana recibir o libro.
c. *É necesario que Ana recibir o libro.
d. *É necesario Ana reciba o libro.
"It is necessary that Ana receives the book."

(26) a. É importante que Marta revise o comunicado.
b. É importante Marta revisar o comunicado.
c. *É importante que Marta revisar o comunicado.
d. *É importante Marta revise o comunicado.
"It is important that Marta revises the statement."

(27) a. É posible que Laura acepte a proposta.
b. É posible Laura aceptar a proposta.
c. *É posible que Laura aceptar a proposta.
d. *É posible Laura acepte a proposta.
"It is possible that Laura accepts the proposal."

(28) a. É probable que Francisco consulte a lei.
b. É probable Francisco consultar a lei.
c. *É probable que Francisco consultar a lei.
d. *É probable Francisco consulte a lei.
"It is probable that Francisco checks the law."

(29) a. É sorprendente que Victoria lea a carta.
b. É sorprendente Victoria ler a carta.
c. *É sorprendente que Victoria ler a carta.
d. *É sorprendente Victoria lea a carta.
"It is surprising that Victoria reads the letter."

(30) a. É necesario que Juan redacte o informe.
b. É necesario Juan redactar o informe.
c. *É necesario que Juan redactar o informe.
d. *É necesario Juan redacte o informe.
“It is necessary that Juan writes the report.”

(31) a. É importante que Salvador contemple o xardín.
b. É importante Salvador contemplar o xardín.
c. *É importante que Salvador contemplar o xardín.
d. *É importante Salvador contemple o xardín.
“It is important that Salvador contemplates the garden.”

(32) a. É posible que Alejandro reforme o almacén.
b. É posible Alejandro reformar o almacén.
c. *É posible que Alejandro reformar o almacén.
d. *É posible Alejandro reforme o almacén.
“It is possible that Alejandro renovates the warehouse.”

(33) a. É probable que Elena pinte o apartamento.
b. É probable Elena pintar o apartamento.
c. *É probable que Elena pintar o apartamento.
d. *É probable Elena pinte o apartamento.
“It is probable that Elena paints the apartment.”

(34) a. É sorprendente que Javier limpe a casa.
b. É sorprendente Javier limpar a casa.
c. *É sorprendente que Javier limpar a casa.
d. *É sorprendente Javier limpe a casa.
“It is surprising that Javier cleans the house.”

(35) a. É necesario que Sara ordene a habitación.
b. É necesario Sara ordenar a habitación.
c. *É necesario que Sara ordenar a habitación.
d. *É necesario Sara ordene a habitación.
“It is necessary that Sara tidies up the room.”

(36) a. É importante que Alba fotografe a escultura.
b. É importante Alba fotografar a escultura.
c. *É importante que Alba fotografar a escultura.
d. *É importante Alba fotografe a escultura.
“It is important that Alba photographs the sculpture.”

(37) a. É posible que Ángela recolla a cociña.
b. É posible Ángela recoller a cociña.
c. *É posible que Ángela recoller a cociña.
d. *É posible Ángela recolla a cociña.
"It is possible that Ángela cleans up the kitchen."

(38) a. É probable que Carlos describa o lago.
b. É probable Carlos describir o lago.
c. *É probable que Carlos describir o lago.
d. *É probable Carlos describa o lago.
"It is probable that Carlos describes the lake."

(39) a. É sorprendente que Julia observe a paisaxe.
b. É sorprendente Julia observar a paisaxe.
c. *É sorprendente que Julia observar a paisaxe.
d. *É sorprendente Julia observe a paisaxe.
"It is surprising that Julia observes the landscape."

(40) a. É necesario que Miguel admire o cadro.
b. É necesario Miguel admirar o cadro.
c. *É necesario que Miguel admirar o cadro.
d. *É necesario Miguel admire o cadro.
"It is necessary that Miguel admires the painting."

(41) a. É importante que Ana modifique a lei.
b. É importante Ana modificar a lei.
c. *É importante que Ana modificar a lei.
d. *É importante Ana modifique a lei.
"It is important that Ana modifies the law."

(42) a. É posible que Manuel reciba a carta.
b. É posible Manuel recibir a carta.
c. *É posible que Manuel recibir a carta.
d. *É posible Manuel reciba a carta.
"It is possible that Manuel receives the letter."

(43) a. É probable que Antonio corrixa o libro.
b. É probable Antonio corrixir o libro.
c. *É probable que Antonio corrixir o libro.
d. *É probable Antonio corrixa o libro.
"It is probable that Antonio corrects the book."

(44) a. É sorprendente que Sofía escriba a postal.
b. É sorprendente Sofía escribir a postal.
c. *É sorprendente que Sofía escribir a postal.
d. *É sorprendente Sofía escriba a postal.
"It is surprising that Sofía writes the postcard."

(45) a. É necesario que José firme o correo.
b. É necesario José firmar o correo.
c. *É necesario que José firmar o correo.
d. *É necesario José firme o correo.
"It is necessary that José signs the email."

(46) a. É importante que Juan lea a novela.
b. É importante Juan ler a novela.
c. *É importante que Juan ler a novela.
d. *É importante Juan lea a novela.
"It is important that Juan reads the novel."

(47) a. É posible que Marta redacte a factura.
b. É posible Marta redactar a factura.
c. *É posible que Marta redactar a factura.
d. *É posible Marta redacte a factura.
"It is possible that Marta writes the invoice."

(48) a. É probable que Laura revise o informe.
b. É probable Laura revisar o informe.
c. *É probable que Laura revisar o informe.
d. *É probable Laura revise o informe.
"It is probable that Laura revises the report."

(49) a. É sorprendente que Francisco acepte o comunicado.
b. É sorprendente Francisco aceptar o comunicado.
c. *É sorprendente que Francisco aceptar o comunicado.
d. *É sorprendente Francisco acepte o comunicado.
"It is surprising that Francisco accepts the statement."

(50) a. É necesario que Victoria consulte a proposta.
b. É necesario Victoria consultar a proposta.
c. *É necesario que Victoria consultar a proposta.
d. *É necesario Victoria consulte a proposta.
"It is necessary that Victoria checks the proposal."

(51) a. É importante que Sara limpe o apartamento.
b. É importante Sara limpar o apartamento.
c. *É importante que Sara limpar o apartamento.
d. *É importante Sara limpe o apartamento.
"It is important that Sara cleans the apartment."

(52) a. É posible que Salvador ordene a casa.
b. É posible Salvador ordenar a casa.
c. *É posible que Salvador ordenar a casa.
d. *É posible Salvador ordene a casa.
"It is possible that Salvador tidies up the house."

(53) a. É probable que Alejandro contemple a escultura.
b. É probable Alejandro contemplar a escultura.
c. *É probable que Alejandro contemplar a escultura.
d. *É probable Alejandro contemple a escultura.
"It is probable that Alejandro contemplates the sculpture."

(54) a. É sorprendente que Elena reforme a cociña.
b. É sorprendente Elena reformar a cociña.
c. *É sorprendente que Elena reformar a cociña.
d. *É sorprendente Elena reforme a cociña.
"It is surprising that Elena renovates the kitchen."

(55) a. É necesario que Javier pinte o almacén.
b. É necesario Javier pintar o almacén.
c. *É necesario que Javier pintar o almacén.
d. *É necesario Javier pinte o almacén.
"It is necessary that Javier paints the warehouse."

(56) a. É importante que Miguel observe o lago.
b. É importante Miguel observar o lago.
c. *É importante que Miguel observar o lago.
d. *É importante Miguel observe o lago.
"It is important that Miguel observes the lake."

(57) a. É posible que Alba admire a paisaxe.
b. É posible Alba admirar a paisaxe.
c. *É posible que Alba admirar a paisaxe.
d. *É posible Alba admire a paisaxe.
"It is possible that Alba admires the landscape."

(58) a. É probable que Ángela fotografe o cadro.
b. É probable Ángela fotografar o cadro.
c. *É probable que Ángela fotografar o cadro.
d. *É probable Ángela fotografe o cadro.
"It is probable that Ángela photographs the painting."

(59) a. É sorprendente que Carlos recolla a habitación.
b. É sorprendente Carlos recoller a habitación.
c. *É sorprendente que Carlos recoller a habitación.
d. *É sorprendente Carlos recolla a habitación.
"It is surprising that Carlos cleans up the room."

(60) a. É necesario que Julia describa o xardín.
b. É necesario Julia describir o xardín.
c. *É necesario que Julia describir o xardín.
d. *É necesario Julia describa o xardín.
"It is necessary that Julia describes the garden."

(61) a. É importante que José escriba o libro.
b. É importante José escribir o libro.
c. *É importante que José escribir o libro.
d. *É importante José escriba o libro.
"It is important that José writes the book."

(62) a. É posible que Ana firme a postal.
b. É posible Ana firmar a postal.
c. *É posible que Ana firmar a postal.
d. *É posible Ana firme a postal.
"It is possible that Ana signs the postcard."

(63) a. É probable que Manuel modifique a proposta.
b. É probable Manuel modificar a proposta.
c. *É probable que Manuel modificar a proposta.
d. *É probable Manuel modifique a proposta.
"It is probable that Manuel modifies the proposal."

(64) a. É sorprendente que Antonio reciba a novela.
b. É sorprendente Antonio recibir a novela.
c. *É sorprendente que Antonio recibir a novela.
d. *É sorprendente Antonio reciba a novela.
"It is surprising that Antonio receives the novel."

(65) a. É necesario que Sofía corrixa a carta.
b. É necesario Sofía corrixir a carta.
c. *É necesario que Sofía corrixir a carta.
d. *É necesario Sofía corrixa a carta.
"It is necessary that Sofía corrects the letter."

(66) a. É importante que Victoria acepte o informe.
b. É importante Victoria aceptar o informe.
c. *É importante que Victoria aceptar o informe.
d. *É importante Victoria acepte o informe.
"It is important that Victoria accepts the report."

(67) a. É posible que Juan consulte o comunicado.
b. É posible Juan consultar o comunicado.
c. *É posible que Juan consultar o comunicado.
d. *É posible Juan consulte o comunicado.
"It is possible that Juan checks the statement."

(68) a. É probable que Marta lea o correo.
b. É probable Marta ler o correo.
c. *É probable que Marta ler o correo.
d. *É probable Marta lea o correo.
"It is probable that Marta reads the email."

(69) a. É importante que Laura redacte a lei.
b. É importante Laura redactar a lei.
c. *É importante que Laura redactar a lei.
d. *É importante Laura redacte a lei.
"It is important that Laura writes the law."

(70) a. É necesario que Francisco revise a factura.
b. É necesario Francisco revisar a factura.
c. *É necesario que Francisco revisar a factura.
d. *É necesario Francisco revise a factura.
"It is necessary that Francisco revises the invoice."

(71) a. É importante que Javier reforme a habitación.
b. É importante Javier reformar a habitación.
c. *É importante que Javier reformar a habitación.
d. *É importante Javier reforme a habitación.
"It is important that Javier renovates the room."

(72) a. É posible que Sara pinte a cociña.
b. É posible Sara pintar a cociña.
c. *É posible que Sara pintar a cociña.
d. *É posible Sara pinte a cociña.
"It is possible that Sara paints the kitchen."

(73) a. É probable que Salvador limpe o almacén.
b. É probable Salvador limpar o almacén.
c. *É probable que Salvador limpar o almacén.
d. *É probable Salvador limpe o almacén.
"It is probable that Salvador cleans the warehouse."

(74) a. É sorprendente que Alejandro ordene o apartamento.
b. É sorprendente Alejandro ordenar o apartamento.
c. *É sorprendente que Alejandro ordenar o apartamento.
d. *É sorprendente Alejandro ordene o apartamento.
"It is surprising that Alejandro tidies up the apartment."

(75) a. É necesario que Elena contemple o cadro.
b. É necesario Elena contemplar o cadro.
c. *É necesario que Elena contemplar o cadro.
d. *É necesario Elena contemple o cadro.
"It is necessary that Elena contemplates the painting."

(76) a. É importante que Julia recolla a casa.
b. É importante Julia recoller a casa.
c. *É importante que Julia recoller a casa.
d. *É importante Julia recolla a casa.
"It is important that Julia cleans up the house."

(77) a. É posible que Miguel describa a escultura.
b. É posible Miguel describir a escultura.
c. *É posible que Miguel describir a escultura.
d. *É posible Miguel describa a escultura.
"It is possible that Miguel describes the sculpture."

(78) a. É probable que Alba observe o xardín.
b. É probable Alba observar o xardín.
c. *É probable que Alba observar o xardín.
d. *É probable Alba observe o xardín.
"It is probable that Alba observes the garden."

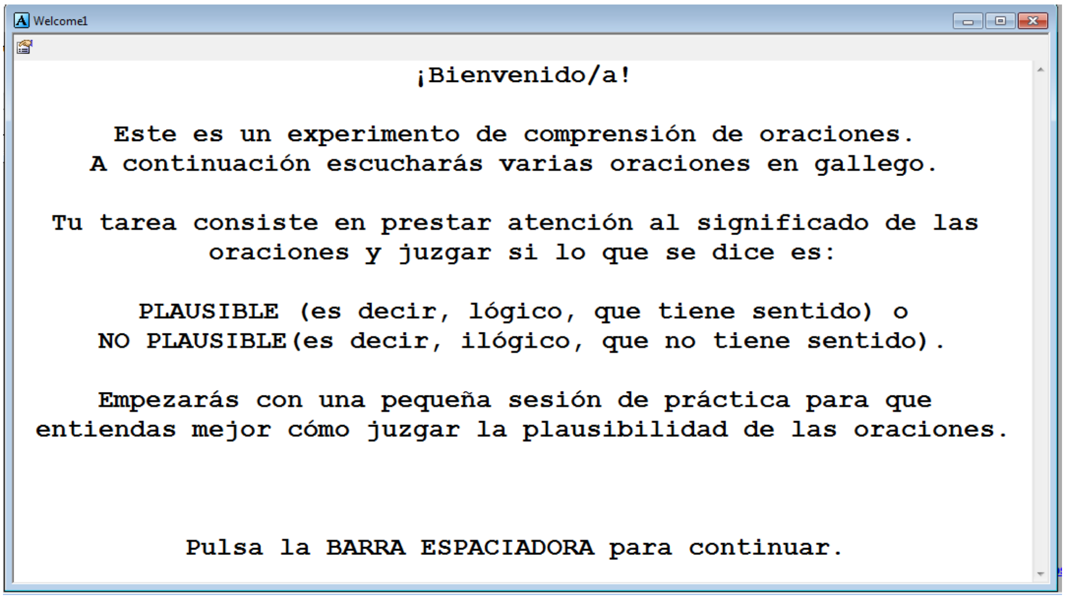
(79) a. É sorprendente que Ángela admire o lago.
b. É sorprendente Ángela admirar o lago.
c. *É sorprendente que Ángela admirar o lago.
d. *É sorprendente Ángela admire o lago.
"It is surprising that Ángela admires the lake."

(80) a. É necesario que Carlos fotografe a paisaxe.
b. É necesario Carlos fotografar a paisaxe.
c. *É necesario que Carlos fotografar a paisaxe.
d. *É necesario Carlos fotografe a paisaxe.
"It is necessary that Carlos signs the landscape."

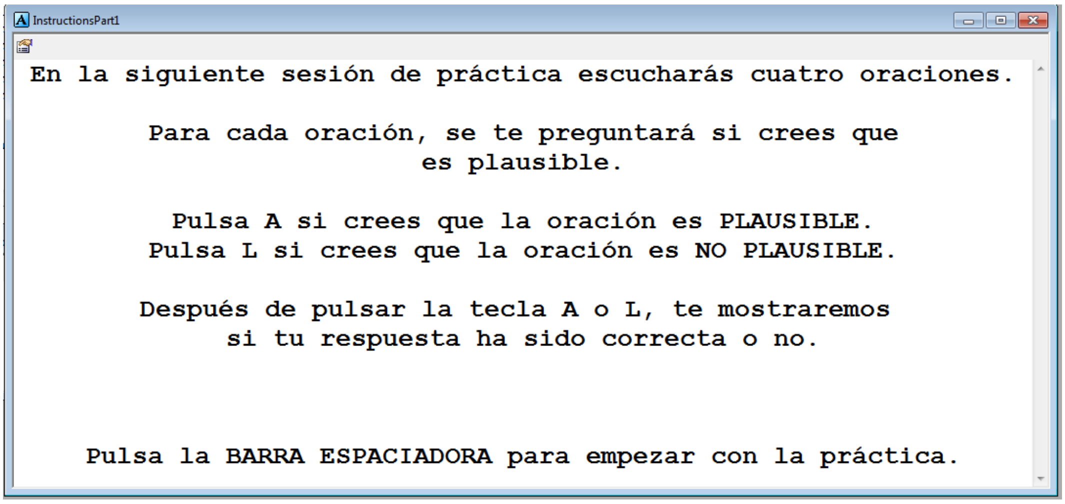
7. Instructions used in Experiment 1

This appendix includes the instructions for the exposure phase, the testing phase and the debriefing phase in Experiment 1. Instructions are presented in Spanish, the language of the experiment. The English translation is presented below each slide.

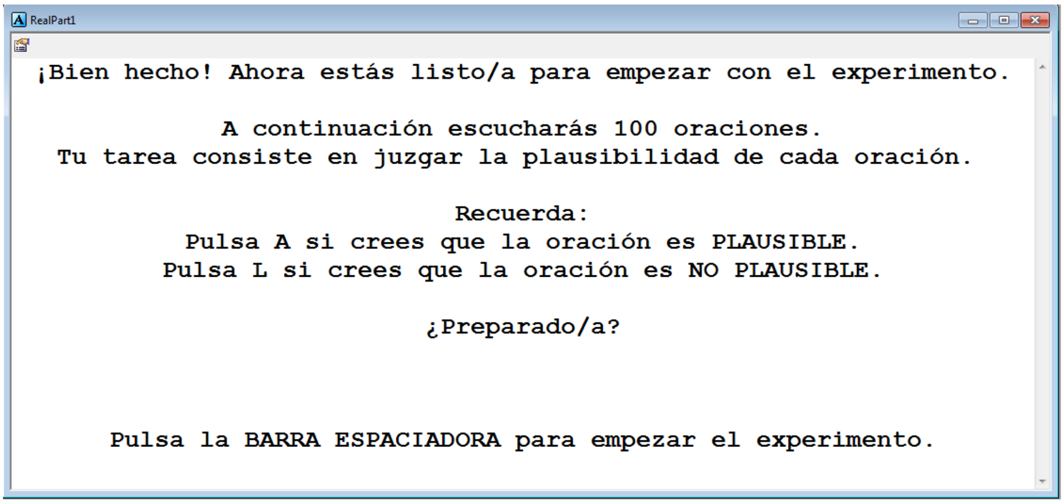
<u>Exposure phase</u>



ENG: Welcome! This is an experiment on sentence comprehension. You will listen to some sentences in Galician. Your task is to pay attention to the meaning of the sentences and to judge whether they are: PLAUSIBLE (that is, logical, that makes sense) or NOT PLAUSIBLE (that is, illogical, that does not make sense). You will begin with a short practice session that will help you better understand how to make plausibility judgements. Press the SPACE BAR to continue.

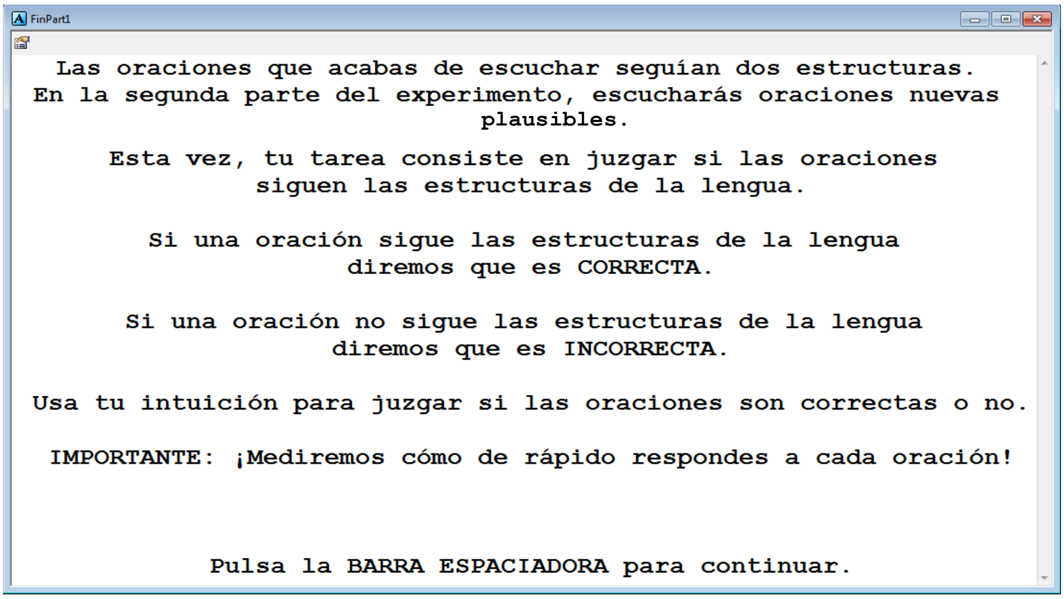


ENG: In the following practice session, you will listen to four sentences. For each sentence, you will be asked whether it is plausible. Press A if you think that the sentence is PLAUSIBLE. Press L if you think that the sentence is NOT PLAUSIBLE. After pressing A or L, we will tell you whether your answer was correct or not. Press the SPACE BAR to begin with the practice.

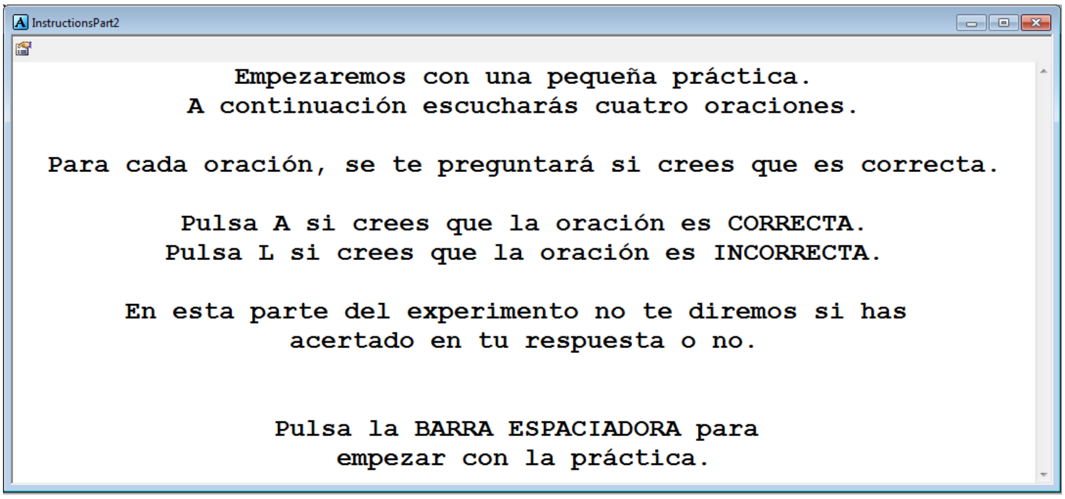


ENG: Well done! You are now ready to begin with the experiment. You will listen to a hundred sentences. Your task is to judge the plausibility of each sentence. Remember: Press A if you think that the sentence is PLAUSIBLE. Press L if you think that the sentence is NOT PLAUSIBLE. Ready? Press the SPACE BAR to begin with the experiment.

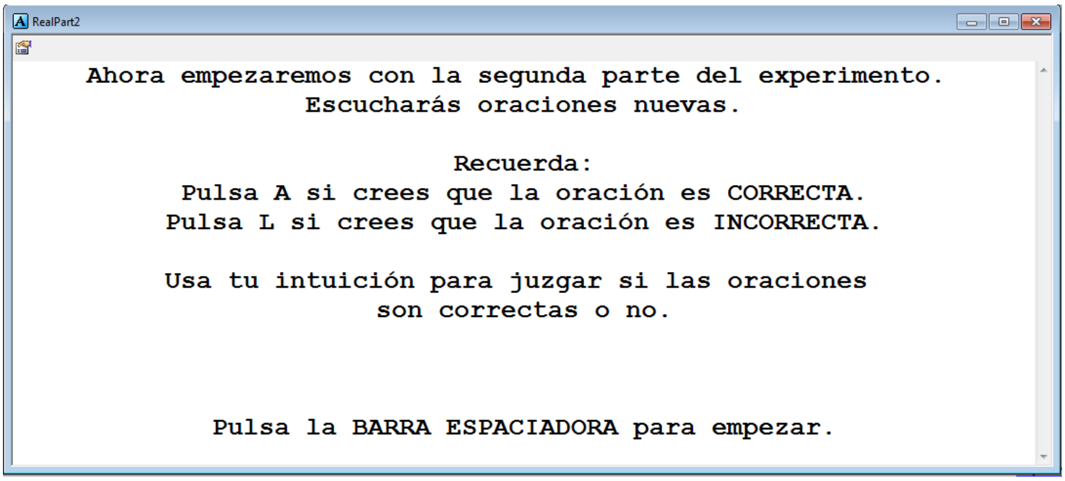
<u>Testing phase</u>



ENG: The sentences you just listened to were formed according to two structures. In the second part of the experiment, you will listen to new plausible sentences. This time, your task is to judge whether the sentences are formed according to the structures of the language. If a sentence is formed according to the structures of the language, we say it is CORRECT. If a sentence is not formed according to the structures of the language, we say it is INCORRECT. Use your intuition to judge whether the sentences are correct or not. IMPORTANT: We will measure your response times! Press the SPACE BAR to continue.

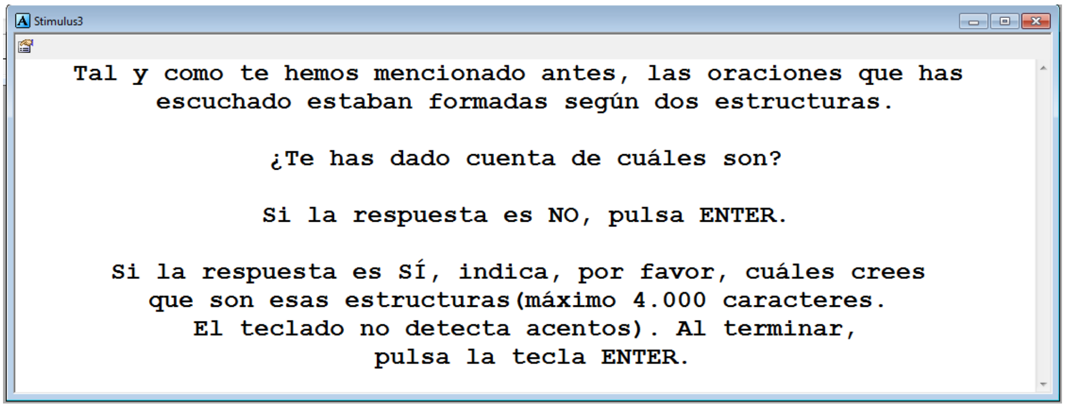


ENG: You will begin with a short practice. You will listen to four sentences. For each sentence, you will be asked whether it is correct. Press A if you think that the sentence is CORRECT. Press L if you think that the sentence is INCORRECT. In this part of the experiment, we will not tell you whether your answer was right. Press the SPACE BAR to begin with the practice.

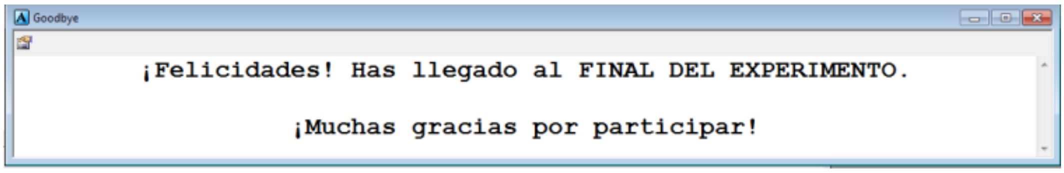


ENG: You will now start with the second part of the experiment. You will listen to new sentences. Remember: Press A if you think that the sentence is CORRECT. Press L if you think that the sentence is INCORRECT. Use your intuition to judge whether the sentences are correct or not. Press the SPACE BAR to begin.

<u>Debriefing phase</u>



ENG: As previously mentioned, the sentences you listened to were formed according to two structures. Did you notice which structures were these? If your answer is NO, press ENTER. If your answer is YES, indicate, please, which you think these structures are (maximum 4,000 characters. The keyboard does not admit accent marks). When you finish, press ENTER.



ENG: Congratulations! You have reached the END OF THE EXPERIMENT. Thank you for participating!

8. Rubric to evaluate awareness and transcription of verbal reports in Experiments 1 and 2

<u>Rubric evaluating awareness (Experiments 1 and 2)</u>

The questions asked in the verbal report were:

1. As mentioned during the experiment, the sentences you listened to were formed according to two structures. Did you notice which structures were these?
2. If yes, please indicate which you think these structures are.

I present below the rubric used to evaluate awareness of the structures based on answers to these questions.

Participant status	Description
Aware	The participant states that a structure contains the complementizer *que* and then a verb conjugated (in the present subjunctive). Additionally or alternatively, the participant reports that the other structure does not contain the complementizer *que* and contains a verb in the infinitive.
Unaware	The participant is not able to identify the varying elements in the structures, i.e. the presence/absence of the complementizer and the finite or non-finite nature of the embedded verb. Alternatively, s/he is able to identify them but cannot appropriately correlate them.

TABLE A-8.1. Rubric used to classify participants as *aware* or *unaware* of the L2 structures based on their responses in the verbal report of Experiments 1 and 2.

<u>Verbal reports and awareness (Experiment 1)</u>

Participant	Report	Awareness
P1	Se trata de oraciones subjuntivas con *que*. Una parte está en presente simple del subjuntivo y otra en pretérito simple.	Aware
P2	-	Unaware
P3	-	Unaware
P4	Al escuchar las diferentes frases me he dado cuenta de que unas contenían la palabra *que* y las otras no.	Unaware
P5	-	Unaware
P6	-	Unaware
P7	-	Unaware

Participant	**Report**	**Awareness**
P8	Las oraciones siguen una estructura que empieza por *Es*, posteriormente un verbo como *importante* o *necesario* y después otro verbo.	Unaware
P9	Utilizar *que* antes del sujeto o no.	Unaware
P10	El nexo de unión *que*. Comerse o no comerse palabras.	Unaware
P11	-	Unaware
P12	-	Unaware
P13	Es el uso del *que* frente al nombre.	Unaware
P14	-	Unaware
P15	-	Unaware
P16	-	Unaware
P17	-	Unaware
P18	En algunas oraciones el nombre va en medio, mientras que en otras cambia de sitio al principio o al final.	Unaware
P19	El *que* antes del nombre de la persona para referirse a quien lo hace.	Unaware
P20	En algunas frases está el nexo *que* y en otras no.	Unaware
P21	Algunas frases contienen la palabra *que* y otras no.	Unaware
P22	-	Unaware
P23	-	Unaware
P24	La presencia o no del *que* y la terminación del verbo.	Unaware

TABLE A-8.2. Transcription of participants' responses in the verbal report of Experiment 1 (in Spanish). Based on these responses, participants are classified as *aware* or *unaware* of the L2 structures.

<u>English translation of verbal reports</u>

P1: These are subjunctive sentences with *that*. Some are in the present subjunctive and others in the past simple.

P4: When listening to the different sentences I realized that some contained the word *that* and others did not.

P8: The sentences follow a structure that begins with *It is*, then a verb such as *important* or *necessary* and then another verb.

P9: Using *that* before the subject or not.

P10: The connecting link *that*. Eating or not eating words.

P13: Using *that* in front of the [proper] noun.

P18: On some occasions, the proper noun appears in the middle of the sentence, while in others it appears at the beginning or at the end.

P19: The word *that* before the person's name is used to refer to the person who does it [the action].

P20: In some sentences there is the word *that* and in others there is not.
P21: Some sentences contain the word *that* and others do not.
P24: The presence or absence of *that* and the ending of the verb.

Verbal reports and awareness (Experiment 2)

Participant	Report	Awareness
P1	Una de las estructuras se componía de la palabra *que*, más quien realiza la acción, más un verbo conjugado, mientras que la otra estructura carecía de *que* y luego tenía la persona que realiza la acción y un verbo en infinitivo.	Aware
P2	Las oraciones que no presentan un relativo *que* presentan el verbo en infinitivo, mientras que las que presentan dicho pronombre presentan un verbo conjugado.	Aware
P3	Las estructuras que incluyen *que* y las que no.	Unaware
P4	Unas oraciones estaban hechas con *que* más subjuntivo y las otras oraciones tenían directamente el infinitivo.	Aware
P5	En las oraciones subordinadas con *que* el verbo que le sigue tiene que estar en subjuntivo, pero si la oración subordinada no tiene *que,* el verbo debe estar en infinitivo.	Aware
P6	-	Unaware
P7	Cuando la oración lleva la palabra *que* el verbo aparece conjugado, mientras que si la oración no lleva la palabra *que* delante del nombre de la persona, el verbo aparece en infinitivo.	Aware
P8	En las oraciones hay una serie de verbos impersonales que rigen el *que* y otros verbos impersonales que no lo presentan o no lo requieren.	Unaware
P9	Cuando el verbo va seguido de *que,* el siguiente verbo tiene que estar conjugado. Cuando no va seguido de *que,* el siguiente verbo no tiene que presentar conjugación.	Aware
P10	Cuando en la frase está el complemento *que*, el verbo no debe ir en infinitivo, ej. oración correcta: *…que acabe…*; ej. oración incorrecta: *…que acabar*. Si no presenta *que* la oración, debe ir en infinitivo el verbo, ej: *…acabar…*	Aware
P11	-	Unaware
P12	En castellano es como si dijésemos > *Es fácil que Andrea haga algo — que* más verbo conjugado. El otro tipo de oración sería > *Es fácil Andrea hacer algo* —no escribimos *que* y el verbo en lugar de conjugado se encuentra en infinitivo.	Aware
P13	-	Unaware
P14	Cuando detrás del primer verbo hay un *que,* el segundo verbo está conjugado. Cuando no hay *que,* el segundo verbo se queda en infinitivo.	Aware

Participant	**Report**	**Awareness**
P15	Cuando aparece el *que* en la frase, el verbo se conjuga. Cuando no aparece el *que,* el verbo se queda en infinitivo.	Aware
P16	Si la oración tiene *que,* el verbo se conjuga. Si la oración no tiene *que,* el verbo aparece en infinitivo.	Aware
P17	El *que* (o no *que*) tiene algo que ver con la variación de las frases.	Unaware
P18	En las oraciones que contienen la palabra *que,* el verbo va conjugado. En cambio, en las frases en las que no aparece el *que,* el verbo va en infinitivo.	Aware
P19	La primera estructura es: *É importante, necesario… que* más verbo conjugado.	Aware
P20	Una de ellas tiene la misma estructura que el castellano: *Es importante que Amaia compre el pan.* Es decir, el verbo está conjugado. Sin embargo, en la otra estructura desaparece el *que* y el verbo aparece en infinitivo. Por lo tanto, la oración que he puesto de ejemplo quedaría en castellano de la siguiente manera: *Es importante Amaia comprar pan*.	Aware
P21	-	Unaware
P22	Una de las estructuras tenía una oración subordinada con el verbo en infinitivo y la otra estructura introducía la frase subordinada con un *que* y el verbo en modo subjuntivo.	Aware
P23	-	Unaware
P24	En las frases con *que* el verbo está conjugado y en las que no hay *que* el verbo está en infinitivo.	Aware
P25	Si la frase tenía *que,* el verbo iba conjugado. Si la frase no tenía *que*, el verbo iba sin conjugar. Es decir, terminaba en vocal + *r*.	Aware
P26	Por un lado, hay un tipo de estructura que está formada por el verbo en infinitivo simplemente. Por otro lado, en el otro tipo de estructura las oraciones estaban formadas por *que* + el verbo conjugado.	Aware
P27	Si la oración lleva *que,* el verbo no va en infinitivo. Si la oración no lleva *que,* el verbo irá en infinitivo.	Aware
P28	Las dos eran oraciones subordinadas sustantivas, pero una estructura era con infinitivo y la otra con *que* más el verbo en subjuntivo.	Aware
P29	-	Unaware
P30	Las oraciones con *que* usan un subjuntivo, mientras que las frases sin el *que* usan infinitivo.	Aware
P31	-	Unaware
P32	Estructura 1: *É necesario que* + sujeto + verbo conjugado + complementos.	Aware

Participant	Report	Awareness
P33	Unas oraciones eran subordinadas y las otras eran impersonales. Las subordinadas usaban verbos en subjuntivo y las impersonales usaban verbos en infinitivo.	Unaware
P34	En una de las estructuras está la conjunción *que* y el verbo aparece conjugado. En la otra estructura no aparece la conjunción *que* y el verbo aparece en infinitivo.	Aware
P35	En las oraciones donde hay un nexo como en este caso es el *que*, se necesita una conjugación del verbo para complementar la oración de manera correcta.	Aware
P36	Unas oraciones tienen el relativo *que* con el verbo en subjuntivo y en las otras no hay *que* y el verbo está en infinitivo.	Aware
P37	Existen dos estructuras de subordinación, la primera de ellas introduce la oración subordinada mediante el nexo *que*, mientras que en la segunda se introduce directamente mediante el verbo en infinitivo.	Aware
P38	*Por* y *para*. Cuando el verbo se conjuga, es *por* y cuando no se conjuga, *para*.	Unaware
P39	La primera estructura está formada por el verbo *ser*, un adjetivo como *probable, necesario…* y una oración subordinada de complemento directo en la que el verbo está en subjuntivo. Esta estructura es la misma que en castellano. La segunda estructura está formada por el mismo comienzo, pero sin utilizar *que* para introducir la subordinada y el verbo va en infinitivo.	Aware
P40	-	Unaware
P41	-	Unaware
P42	En algunas oraciones hay un *que* y en otras no. No depende de lo que va antes (es decir, *é posible…*), puede ir *que* o no. Probablemente depende del tipo de verbo que sigue después.	Unaware
P43	-	Unaware
P44	Las oraciones que tienen *que* van seguidas de subjuntivo y las que no tienen ese *que* antes del nombre de la persona, van seguidas del infinitivo.	Aware

TABLE A-8.3. Transcription of participants' responses in the verbal report of Experiment 2 (in Spanish). Based on these responses, participants are classified as *aware* or *unaware* of the L2 structures.

English translation of verbal reports

P1: One of the structures consisted of the word *that*, plus the person performing the action, plus a conjugated verb. The other structure lacked *that* and then had the person performing the action and an infinitive verb.

P2: Sentences that do not include a relative *that* have the verb in the infinitive, while those that include the pronoun have a conjugated verb.

P3: The structures that include *that* and those that do not.

P4: Some sentences contained *that* plus subjunctive and other sentences had just an infinitive.

P5: In subordinate clauses with *that* the verb that follows must be in the subjunctive, but if the subordinate clause does not have *that,* the verb must be in the infinitive.

P7: When the sentence contains the word *that,* the verb appears conjugated, while if the sentence does not contain the word *that* in front of the person's name, the verb appears in the infinitive.

P8: In the sentences, there are a series of impersonal verbs governing *that* and other impersonal verbs that do not require it.

P9: When a verb is followed by *that,* the following verb must be conjugated. When it is not followed by *that,* the following verb is not conjugated.

P10: When the complement *that* is present in the sentence, the verb must not be in the infinitive. If the sentence does not have *that*, the verb must be in the infinitive.

P12: In Spanish, it is as if we were saying > *It is easy that Andrea does something* – *that* plus conjugated verb. The other type of sentence would be > *It is easy Andrea to do something* – we do not write *that* and the verb, instead of being conjugated, is in the infinitive.

P14: When after the first verb there is *that,* the second verb is conjugated. When there is no *that,* the second verb remains in the infinitive.

P15: When *that* appears in the sentence, the verb is conjugated. When *that* does not appear, the verb remains in the infinitive.

P16: If the sentence has *that*, the verb is conjugated. If the sentence does not have *that*, the verb appears in the infinitive.

P17: The [presence of] *that* (or no *that*) has something to do with the variation in the sentences.

P18: In sentences that contain the word *that*, the verb is conjugated. On the other hand, in sentences in which *that* does not appear, the verb is in the infinitive.

P19: The first structure is: *It is important, necessary*… *that* plus conjugated verb.

P20: One of them has the same structure as in Spanish: *It is important that Amaia buys the bread*. That is, the verb is conjugated. However, in the other structure *that* disappears and the verb appears in the infinitive. Therefore, the sentence that I have given as an example would be in Spanish as follows: *It is important Amaia to buy bread*.

P22: One of the structures had a subordinate clause with the verb in the infinitive and the other structure introduced the subordinate clause with *that* and had the verb in the subjunctive mood.

P24: In sentences with *that*, the verb is conjugated and in those without *that*, the verb is in the infinitive.

P25: If the sentence had *that*, the verb was conjugated. If the sentence did not have *that*, the verb was not conjugated, that is, it ended in a vowel + *r*.

P26: On the one hand, there is a type of structure that is formed by the verb in the infinitive. On the other hand, the other type of structure is formed by *that* + a conjugated verb.

P27: If the sentence has *that,* the verb is not in the infinitive. If the sentence does not have *that,* the verb is in the infinitive.

P28: Both were substantive subordinate clauses, but one structure had an infinitive and the other had *that* plus the verb in the subjunctive.

P30: Sentences with *that* have a subjunctive, while sentences without *that* have an infinitive.

P32: Structure 1: *It is necessary that* + subject + conjugated verb + complements.

P33: Some sentences were subordinate and others were impersonal. The subordinate ones used subjunctive verbs and the impersonal ones used infinitive verbs.

P34: In one of the structures there is the conjunction *that* and the verb appears conjugated. In the other structure the conjunction *that* does not appear and the verb appears in the infinitive.

P35: In sentences where there is a link, such as *that*, conjugating the verb is necessary to form the sentence correctly.

P36: Some sentences have the relative *that* and the verb in the subjunctive and in others there is no *that* and the verb is in the infinitive.

P37: There are two subordination structures. The first of them introduces the subordinate clause with the link *that* while, in the second, this is introduced directly with the infinitive verb.

P38: *By* and *for*. When the verb is conjugated, it is *by* and when it is not conjugated, it is *for*.

P39: The first structure is made up of the verb *to be*, an adjective such as *probable, necessary…* and a subordinate clause in which the verb is in the subjunctive. This structure is the same as in Spanish. The second structure is formed by the same beginning, but without using *that* to introduce the subordinate clause and the verb is in the infinitive.

P42: In some sentences, there is *that* and in others, there is not. It does not depend on what goes before (that is, *it is possible…*). It probably depends on the type of verb that follows it.

P44: Sentences that have *that* are followed by the subjunctive and those that do not have *that* before the name of the person are followed by the infinitive.

9. Experimental materials used in Experiment 2

Exposure set (Plausible sentences in Experiment 1)

Galician-based sentences constituting the exposure set in Experiment 2. Sentences are presented in pairs, corresponding to the two conditions in the exposure phase: a. Plausible similar structure and b. Plausible dissimilar structure.

(1) a. É importante que Pedro repare a radio.
b. É importante Pedro reparar a radio.

(2) a. É posible que Mónica apague a consola.
b. É posible Mónica apagar a consola.

(3) a. É probable que Ángel acenda o móbil.
b. É probable Ángel acender o móbil.

(4) a. É sorprendente que Isabel venda o portátil.
b. É sorprendente Isabel vender o portátil.

(5) a. É necesario que Pablo compre o televisor.
b. É necesario Pablo comprar o televisor.

(6) a. É importante que Lucía saboree a torta.
b. É importante Lucía saborear a torta.

(7) a. É posible que Luis corte a carne.
b. É posible Luis cortar a carne.

(8) a. É probable que Paula cociñe o pan.
b. É probable Paula cociñar o pan.

(9) a. É sorprendente que Jorge mastigue o brócoli.
b. É sorprendente Jorge mastigar o brócoli.

(10) a. É necesario que Alicia conxele a cenoria.
b. É necesario Alicia conxelar a cenoria.

(11) a. É importante que Alberto gañe o concurso.
b. É importante Alberto gañar o concurso.

(12) a. É posible que Silvia perda a proba.
b. É posible Silvia perder a proba.

(13) a. É probable que Adrián financie o torneo.
b. É probable Adrián financiar o torneo.

(14) a. É sorprendente que Irene organice a carreira.
b. É sorprendente Irene organizar a carreira.

(15) a. É necesario que Iván cancele o campionato.
b. É necesario Iván cancelar o campionato.

(16) a. É importante que Rosa devolva a xoia.
b. É importante Rosa devolver a xoia.

(17) a. É posible que Raúl garde a pulseira.
b. É posible Raúl gardar a pulseira.

(18) a. É probable que Andrea encargue o abrigo.
b. É probable Andrea encargar o abrigo.

(19) a. É sorprendente que Óscar envolva o reloxo.
b. É sorprendente Óscar envolver o reloxo.

(20) a. É necesario que Carmen envíe o xersei.
b. É necesario Carmen enviar o xersei.

(21) a. É importante que Pablo venda o móbil.
b. É importante Pablo vender o móbil.

(22) a. É posible que Pedro compre o portátil.
b. É posible Pedro comprar o portátil.

(23) a. É probable que Mónica repare o televisor.
b. É probable Mónica reparar o televisor.

(24) a. É sorprendente que Ángel apague a radio.
b. É sorprendente Ángel apagar a radio.

(25) a. É necesario que Isabel acenda a consola.
b. É necesario Isabel acender a consola.

(26) a. É importante que Alicia mastigue o pan.
b. É importante Alicia mastigar o pan.

(27) a. É posible que Lucía conxele o brócoli.
b. É posible Lucía conxelar o brócoli.

(28) a. É probable que Luis saboree a cenoria.
b. É probable Luis saborear a cenoria.

(29) a. É sorprendente que Paula corte a torta.
b. É sorprendente Paula cortar a torta.

(30) a. É necesario que Jorge cociñe a carne.
b. É necesario Jorge cociñar a carne.

(31) a. É importante que Iván organice o torneo.
b. É importante Iván organizar o torneo.

(32) a. É posible que Alberto cancele a carreira.
b. É posible Alberto cancelar a carreira.

(33) a. É probable que Silvia gañe o campionato.
b. É probable Silvia gañar o campionato.

(34) a. É sorprendente que Adrián perda o concurso.
b. É sorprendente Adrián perder o concurso.

(35) a. É necesario que Irene financie a proba.
b. É necesario Irene financiar a proba.

(36) a. É importante que Carmen envolva o abrigo.
b. É importante Carmen envolver o abrigo.

(37) a. É posible que Rosa envíe o reloxo.
b. É posible Rosa enviar o reloxo.

(38) a. É probable que Raúl devolva o xersei.
b. É probable Raúl devolver o xersei.

(39) a. É sorprendente que Andrea garde a xoia.
b. É sorprendente Andrea gardar a xoia.

(40) a. É necesario que Óscar encargue a pulseira.
b. É necesario Óscar encargar a pulseira.

(41) a. É importante que Isabel apague o televisor.
b. É importante Isabel apagar o televisor.

(42) a. É posible que Pablo acenda a radio.
b. É posible Pablo acender a radio.

(43) a. É probable que Pedro venda a consola.
b. É probable Pedro vender a consola.

(44) a. É sorprendente que Mónica compre o móbil.
b. É sorprendente Mónica comprar o móbil.

(45) a. É necesario que Ángel repare o portátil.
b. É necesario Ángel reparar o portátil.

(46) a. É importante que Jorge corte a cenoria.
b. É importante Jorge cortar a cenoria.

(47) a. É posible que Alicia cociñe a torta.
b. É posible Alicia cociñar a torta.

(48) a. É probable que Lucía mastigue a carne.
b. É probable Lucía mastigar a carne.

(49) a. É sorprendente que Luis conxele o pan.
b. É sorprendente Luis conxelar o pan.

(50) a. É necesario que Paula saboree o brócoli.
b. É necesario Paula saborear o brócoli.

(51) a. É importante que Irene perda o campionato.
b. É importante Irene perder o campionato.

(52) a. É posible que Iván financie o concurso.
b. É posible Iván financiar o concurso.

(53) a. É probable que Alberto organice a proba.
b. É probable Alberto organizar a proba.

(54) a. É sorprendente que Silvia cancele o torneo.
b. É sorprendente Silvia cancelar o torneo.

(55) a. É necesario que Adrián gañe a carreira.
b. É necesario Adrián gañar a carreira.

(56) a. É importante que Óscar garde o xersei.
b. É importante Óscar gardar o xersei.

(57) a. É posible que Carmen encargue a xoia.
b. É posible Carmen encargar a xoia.

(58) a. É probable que Rosa envolva a pulseira.
b. É probable Rosa envolver a pulseira.

(59) a. É sorprendente que Raúl envíe o abrigo.
b. É sorprendente Raúl enviar o abrigo.

(60) a. É necesario que Andrea devolva o reloxo.
b. É necesario Andrea devolver o reloxo.

(61) a. É importante que Ángel compre a consola.
b. É importante Ángel comprar a consola.

(62) a. É posible que Isabel repare o móbil.
b. É posible Isabel reparar o móbil.

(63) a. É probable que Pablo apague o portátil.
b. É probable Pablo apagar o portátil.

(64) a. É sorprendente que Pedro acenda o televisor.
b. É sorprendente Pedro acender o televisor.

(65) a. É necesario que Mónica venda a radio.
b. É necesario Mónica vender a radio.

(66) a. É importante que Paula conxele a carne.
b. É importante Paula conxelar a carne.

(67) a. É posible que Jorge saboree o pan.
b. É posible Jorge saborear o pan.

(68) a. É probable que Alicia corte o brócoli.
b. É probable Alicia cortar o brócoli.

(69) a. É sorprendente que Lucía cociñe a cenoria.
b. É sorprendente Lucía cociñar a cenoria.

(70) a. É necesario que Luis mastigue a torta.
b. É necesario Luis mastigar a torta.

(71) a. É importante que Adrián cancele a proba.
b. É importante Adrián cancelar a proba.

(72) a. É posible que Irene gañe o torneo.
b. É posible Irene gañar o torneo.

(73) a. É probable que Iván perda a carreira.
b. É probable Iván perder a carreira.

(74) a. É sorprendente que Alberto financie o campionato.
b. É sorprendente Alberto financiar o campionato.

(75) a. É necesario que Silvia organice o concurso.
b. É necesario Silvia organizar o concurso.

(76) a. É importante que Andrea envíe a pulseira.
b. É importante Andrea enviar a pulseira.

(77) a. É posible que Óscar devolva o abrigo.
b. É posible Óscar devolver o abrigo.

(78) a. É probable que Carmen garde o reloxo.
b. É probable Carmen gardar o reloxo.

(79) a. É sorprendente que Rosa encargue o xersei.
b. É sorprendente Rosa encargar o xersei.

(80) a. É necesario que Raúl envolva a xoia.
b. É necesario Raúl envolver a xoia.

(81) a. É importante que Mónica acenda o portátil.
b. É importante Mónica acender o portátil.

(82) a. É posible que Ángel venda o televisor.
b. É posible Ángel vender o televisor.

(83) a. É probable que Isabel compre a radio.
b. É probable Isabel comprar a radio.

(84) a. É sorprendente que Pablo repare a consola.
b. É sorprendente Pablo reparar a consola.

(85) a. É necesario que Pedro apague o móbil.
b. É necesario Pedro apagar o móbil.

(86) a. É importante que Luis cociñe o brócoli.
b. É importante Luis cociñar o brócoli.

(87) a. É posible que Paula mastigue a cenoria.
b. É posible Paula mastigar a cenoria.

(88) a. É probable que Jorge conxele a torta.
b. É probable Jorge conxelar a torta.

(89) a. É sorprendente que Alicia saboree a carne.
b. É sorprendente Alicia saborear a carne.

(90) a. É necesario que Lucía corte o pan.
b. É necesario Lucía cortar o pan.

(91) a. É importante que Silvia financie a carreira.
b. É importante Silvia financiar a carreira.

(92) a. É posible que Adrián organice o campionato.
b. É posible Adrián organizar o campionato.

(93) a. É probable que Irene cancele o concurso.
b. É probable Irene cancelar o concurso.

(94) a. É sorprendente que Iván gañe a proba.
b. É sorprendente Iván gañar a proba.

(95) a. É necesario que Alberto perda o torneo.
b. É necesario Alberto perder o torneo.

(96) a. É importante que Raúl encargue o reloxo.
b. É importante Raúl encargar o reloxo.

(97) a. É posible que Andrea envolva o xersei.
b. É posible Andrea envolver o xersei.

(98) a. É probable que Óscar envíe a xoia.
b. É probable Óscar enviar a xoia.

(99) a. É sorprendente que Carmen devolva a pulseira.
b. É sorprendente Carmen devolver a pulseira.

(100) a. É necesario que Rosa garde o abrigo.
b. É necesario Rosa gardar o abrigo.

<u>Testing set (Same as in Experiment 1)</u>

Galician-based sentences constituting the testing set in Experiment 2. Sentences are presented in groups of four, corresponding to the four conditions in the testing phase: a. Grammatical similar structure, b. Grammatical dissimilar structure, c. Ungrammatical similar structure and d. Ungrammatical dissimilar structure.

(1) a. É importante que Antonio firme a carta.
b. É importante Antonio firmar a carta.
c. *É importante que Antonio firmar a carta.
d. *É importante Antonio firme a carta.

(2) a. É posible que Sofía modifique o informe.
b. É posible Sofía modificar o informe.
c. *É posible que Sofía modificar o informe.
d. *É posible Sofía modifique o informe.

(3) a. É probable que José reciba a postal.
b. É probable José recibir a postal.
c. *É probable que José recibir a postal.
d. *É probable José reciba a postal.

(4) a. É sorprendente que Ana corrixa o correo.
b. É sorprendente Ana corrixir o correo.
c. *É sorprendente que Ana corrixir o correo.
d. *É sorprendente Ana corrixa o correo.

(5) a. É necesario que Manuel escriba a novela.
b. É necesario Manuel escribir a novela.
c. *É necesario que Manuel escribir a novela.
d. *É necesario Manuel escriba a novela.

(6) a. É importante que Laura consulte a factura.
b. É importante Laura consultar a factura.
c. *É importante que Laura consultar a factura.
d. *É importante Laura consulte a factura.

(7) a. É posible que Francisco lea o libro.
b. É posible Francisco ler o libro.
c. *É posible que Francisco ler o libro.
d. *É posible Francisco lea o libro.

(8) a. É probable que Victoria redacte o comunicado.
b. É probable Victoria redactar o comunicado.
c. *É probable que Victoria redactar o comunicado.
d. *É probable Victoria redacte o comunicado.

(9) a. É sorprendente que Juan revise a proposta.
b. É sorprendente Juan revisar a proposta.
c. *É sorprendente que Juan revisar a proposta.
d. *É sorprendente Juan revise a proposta.

(10) a. É necesario que Marta acepte a lei.
b. É necesario Marta aceptar a lei.
c. *É necesario que Marta aceptar a lei.
d. *É necesario Marta acepte a lei.

(11) a. É importante que Alejandro pinte a casa.
b. É importante Alejandro pintar a casa.
c. *É importante que Alejandro pintar a casa.
d. *É importante Alejandro pinte a casa.

(12) a. É posible que Elena limpe a habitación.
b. É posible Elena limpar a habitación.
c. *É posible que Elena limpar a habitación.
d. *É posible Elena limpe a habitación.

(13) a. É probable que Javier ordene a cociña.
b. É probable Javier ordenar a cociña.
c. *É probable que Javier ordenar a cociña.
d. *É probable Javier ordene a cociña.

(14) a. É sorprendente que Sara contemple o lago.
b. É sorprendente Sara contemplar o lago.
c. *É sorprendente que Sara contemplar o lago.
d. *É sorprendente Sara contemple o lago.

(15) a. É necesario que Salvador reforme o apartamento.
b. É necesario Salvador reformar o apartamento.
c. *É necesario que Salvador reformar o apartamento.
d. *É necesario Salvador reforme o apartamento.

(16) a. É importante que Ángela describa a paisaxe.
b. É importante Ángela describir a paisaxe.
c. *É importante que Ángela describir a paisaxe.
d. *É importante Ángela describa a paisaxe.

(17) a. É posible que Carlos observe o cadro.
b. É posible Carlos observar o cadro.
c. *É posible que Carlos observar o cadro.
d. *É posible Carlos observe o cadro.

(18) a. É probable que Julia admire a escultura.
b. É probable Julia admirar a escultura.
c. *É probable que Julia admirar a escultura.
d. *É probable Julia admire a escultura.

(19) a. É sorprendente que Miguel fotografe o xardín.
b. É sorprendente Miguel fotografar o xardín.
c. *É sorprendente que Miguel fotografar o xardín.
d. *É sorprendente Miguel fotografe o xardín.

(20) a. É necesario que Alba recolla o almacén.
b. É necesario Alba recoller o almacén.
c. *É necesario que Alba recoller o almacén.
d. *É necesario Alba recolla o almacén.

(21) a. É importante que Manuel corrixa a postal.
b. É importante Manuel corrixir a postal.
c. *É importante que Manuel corrixir a postal.
d. *É importante Manuel corrixa a postal.

(22) a. É posible que Antonio escriba o correo.
b. É posible Antonio escribir o correo.
c. *É posible que Antonio escribir o correo.
d. *É posible Antonio escriba o correo.

(23) a. É probable que Sofía firme a novela.
b. É probable Sofía firmar a novela.
c. *É probable que Sofía firmar a novela.
d. *É probable Sofía firme a novela.

(24) a. É sorprendente que José modifique a factura.
b. É sorprendente José modificar a factura.
c. *É sorprendente que José modificar a factura.
d. *É sorprendente José modifique a factura.

(25) a. É necesario que Ana reciba o libro.
b. É necesario Ana recibir o libro.
c. *É necesario que Ana recibir o libro.
d. *É necesario Ana reciba o libro.

(26) a. É importante que Marta revise o comunicado.
b. É importante Marta revisar o comunicado.
c. *É importante que Marta revisar o comunicado.
d. *É importante Marta revise o comunicado.

(27) a. É posible que Laura acepte a proposta.
b. É posible Laura aceptar a proposta.
c. *É posible que Laura aceptar a proposta.
d. *É posible Laura acepte a proposta.

(28) a. É probable que Francisco consulte a lei.
b. É probable Francisco consultar a lei.
c. *É probable que Francisco consultar a lei.
d. *É probable Francisco consulte a lei.

(29) a. É sorprendente que Victoria lea a carta.
b. É sorprendente Victoria ler a carta.
c. *É sorprendente que Victoria ler a carta.
d. *É sorprendente Victoria lea a carta.

(30) a. É necesario que Juan redacte o informe.
b. É necesario Juan redactar o informe.
c. *É necesario que Juan redactar o informe.
d. *É necesario Juan redacte o informe.

(31) a. É importante que Salvador contemple o xardín.
b. É importante Salvador contemplar o xardín.
c. *É importante que Salvador contemplar o xardín.
d. *É importante Salvador contemple o xardín.

(32) a. É posible que Alejandro reforme o almacén.
b. É posible Alejandro reformar o almacén.
c. *É posible que Alejandro reformar o almacén.
d. *É posible Alejandro reforme o almacén.

(33) a. É probable que Elena pinte o apartamento.
b. É probable Elena pintar o apartamento.
c. *É probable que Elena pintar o apartamento.
d. *É probable Elena pinte o apartamento.

(34) a. É sorprendente que Javier limpe a casa.
b. É sorprendente Javier limpar a casa.
c. *É sorprendente que Javier limpar a casa.
d. *É sorprendente Javier limpe a casa.

(35) a. É necesario que Sara ordene a habitación.
b. É necesario Sara ordenar a habitación.
c. *É necesario que Sara ordenar a habitación.
d. *É necesario Sara ordene a habitación.

(36) a. É importante que Alba fotografe a escultura.
b. É importante Alba fotografar a escultura.
c. *É importante que Alba fotografar a escultura.
d. *É importante Alba fotografe a escultura.

(37) a. É posible que Ángela recolla a cociña.
b. É posible Ángela recoller a cociña.
c. *É posible que Ángela recoller a cociña.
d. *É posible Ángela recolla a cociña.

(38) a. É probable que Carlos describa o lago.
b. É probable Carlos describir o lago.
c. *É probable que Carlos describir o lago.
d. *É probable Carlos describa o lago.

(39) a. É sorprendente que Julia observe a paisaxe.
b. É sorprendente Julia observar a paisaxe.
c. *É sorprendente que Julia observar a paisaxe.
d. *É sorprendente Julia observe a paisaxe.

(40) a. É necesario que Miguel admire o cadro.
b. É necesario Miguel admirar o cadro.
c. *É necesario que Miguel admirar o cadro.
d. *É necesario Miguel admire o cadro.

(41) a. É importante que Ana modifique a lei.
b. É importante Ana modificar a lei.
c. *É importante que Ana modificar a lei.
d. *É importante Ana modifique a lei.

(42) a. É posible que Manuel reciba a carta.
b. É posible Manuel recibir a carta.
c. *É posible que Manuel recibir a carta.
d. *É posible Manuel reciba a carta.

(43) a. É probable que Antonio corrixa o libro.
b. É probable Antonio corrixir o libro.
c. *É probable que Antonio corrixir o libro.
d. *É probable Antonio corrixa o libro.

(44) a. É sorprendente que Sofía escriba a postal.
b. É sorprendente Sofía escribir a postal.
c. *É sorprendente que Sofía escribir a postal.
d. *É sorprendente Sofía escriba a postal.

(45) a. É necesario que José firme o correo.
b. É necesario José firmar o correo.
c. *É necesario que José firmar o correo.
d. *É necesario José firme o correo.

(46) a. É importante que Juan lea a novela.
b. É importante Juan ler a novela.
c. *É importante que Juan ler a novela.
d. *É importante Juan lea a novela.

(47) a. É posible que Marta redacte a factura.
b. É posible Marta redactar a factura.
c. *É posible que Marta redactar a factura.
d. *É posible Marta redacte a factura.

(48) a. É probable que Laura revise o informe.
b. É probable Laura revisar o informe.
c. *É probable que Laura revisar o informe.
d. *É probable Laura revise o informe.

(49) a. É sorprendente que Francisco acepte o comunicado.
b. É sorprendente Francisco aceptar o comunicado.
c. *É sorprendente que Francisco aceptar o comunicado.
d. *É sorprendente Francisco acepte o comunicado.

(50) a. É necesario que Victoria consulte a proposta.
b. É necesario Victoria consultar a proposta.
c. *É necesario que Victoria consultar a proposta.
d. *É necesario Victoria consulte a proposta.

(51) a. É importante que Sara limpe o apartamento.
b. É importante Sara limpar o apartamento.
c. *É importante que Sara limpar o apartamento.
d. *É importante Sara limpe o apartamento.

(52) a. É posible que Salvador ordene a casa.
b. É posible Salvador ordenar a casa.
c. *É posible que Salvador ordenar a casa.
d. *É posible Salvador ordene a casa.

(53) a. É probable que Alejandro contemple a escultura.
b. É probable Alejandro contemplar a escultura.
c. *É probable que Alejandro contemplar a escultura.
d. *É probable Alejandro contemple a escultura.

(54) a. É sorprendente que Elena reforme a cociña.
b. É sorprendente Elena reformar a cociña.
c. *É sorprendente que Elena reformar a cociña.
d. *É sorprendente Elena reforme a cociña.

(55) a. É necesario que Javier pinte o almacén.
b. É necesario Javier pintar o almacén.
c. *É necesario que Javier pintar o almacén.
d. *É necesario Javier pinte o almacén.

(56) a. É importante que Miguel observe o lago.
b. É importante Miguel observar o lago.
c. *É importante que Miguel observar o lago.
d. *É importante Miguel observe o lago.

(57) a. É posible que Alba admire a paisaxe.
b. É posible Alba admirar a paisaxe.
c. *É posible que Alba admirar a paisaxe.
d. *É posible Alba admire a paisaxe.

(58) a. É probable que Ángela fotografe o cadro.
b. É probable Ángela fotografar o cadro.
c. *É probable que Ángela fotografar o cadro.
d. *É probable Ángela fotografe o cadro.

(59) a. É sorprendente que Carlos recolla a habitación.
b. É sorprendente Carlos recoller a habitación.
c. *É sorprendente que Carlos recoller a habitación.
d. *É sorprendente Carlos recolla a habitación.

(60) a. É necesario que Julia describa o xardín.
b. É necesario Julia describir o xardín.
c. *É necesario que Julia describir o xardín.
d. *É necesario Julia describa o xardín.

(61) a. É importante que José escriba o libro.
b. É importante José escribir o libro.
c. *É importante que José escribir o libro.
d. *É importante José escriba o libro.

(62) a. É posible que Ana firme a postal.
b. É posible Ana firmar a postal.
c. *É posible que Ana firmar a postal.
d. *É posible Ana firme a postal.

(63) a. É probable que Manuel modifique a proposta.
b. É probable Manuel modificar a proposta.
c. *É probable que Manuel modificar a proposta.
d. *É probable Manuel modifique a proposta.

(64) a. É sorprendente que Antonio reciba a novela.
b. É sorprendente Antonio recibir a novela.
c. *É sorprendente que Antonio recibir a novela.
d. *É sorprendente Antonio reciba a novela.

(65) a. É necesario que Sofía corrixa a carta.
b. É necesario Sofía corrixir a carta.
c. *É necesario que Sofía corrixir a carta.
d. *É necesario Sofía corrixa a carta.

(66) a. É importante que Victoria acepte o informe.
b. É importante Victoria aceptar o informe.
c. *É importante que Victoria aceptar o informe.
d. *É importante Victoria acepte o informe.

(67) a. É posible que Juan consulte o comunicado.
b. É posible Juan consultar o comunicado.
c. *É posible que Juan consultar o comunicado.
d. *É posible Juan consulte o comunicado.

(68) a. É probable que Marta lea o correo.
b. É probable Marta ler o correo.
c. *É probable que Marta ler o correo.
d. *É probable Marta lea o correo.

(69) a. É importante que Laura redacte a lei.
b. É importante Laura redactar a lei.
c. *É importante que Laura redactar a lei.
d. *É importante Laura redacte a lei.

(70) a. É necesario que Francisco revise a factura.
b. É necesario Francisco revisar a factura.
c. *É necesario que Francisco revisar a factura.
d. *É necesario Francisco revise a factura.

(71) a. É importante que Javier reforme a habitación.
b. É importante Javier reformar a habitación.
c. *É importante que Javier reformar a habitación.
d. *É importante Javier reforme a habitación.

(72) a. É posible que Sara pinte a cociña.
b. É posible Sara pintar a cociña.
c. *É posible que Sara pintar a cociña.
d. *É posible Sara pinte a cociña.

(73) a. É probable que Salvador limpe o almacén.
b. É probable Salvador limpar o almacén.
c. *É probable que Salvador limpar o almacén.
d. *É probable Salvador limpe o almacén.

(74) a. É sorprendente que Alejandro ordene o apartamento.
b. É sorprendente Alejandro ordenar o apartamento.
c. *É sorprendente que Alejandro ordenar o apartamento.
d. *É sorprendente Alejandro ordene o apartamento.

(75) a. É necesario que Elena contemple o cadro.
b. É necesario Elena contemplar o cadro.
c. *É necesario que Elena contemplar o cadro.
d. *É necesario Elena contemple o cadro.

(76) a. É importante que Julia recolla a casa.
b. É importante Julia recoller a casa.
c. *É importante que Julia recoller a casa.
d. *É importante Julia recolla a casa.

(77) a. É posible que Miguel describa a escultura.
b. É posible Miguel describir a escultura.
c. *É posible que Miguel describir a escultura.
d. *É posible Miguel describa a escultura.

(78) a. É probable que Alba observe o xardín.
b. É probable Alba observar o xardín.
c. *É probable que Alba observar o xardín.
d. *É probable Alba observe o xardín.

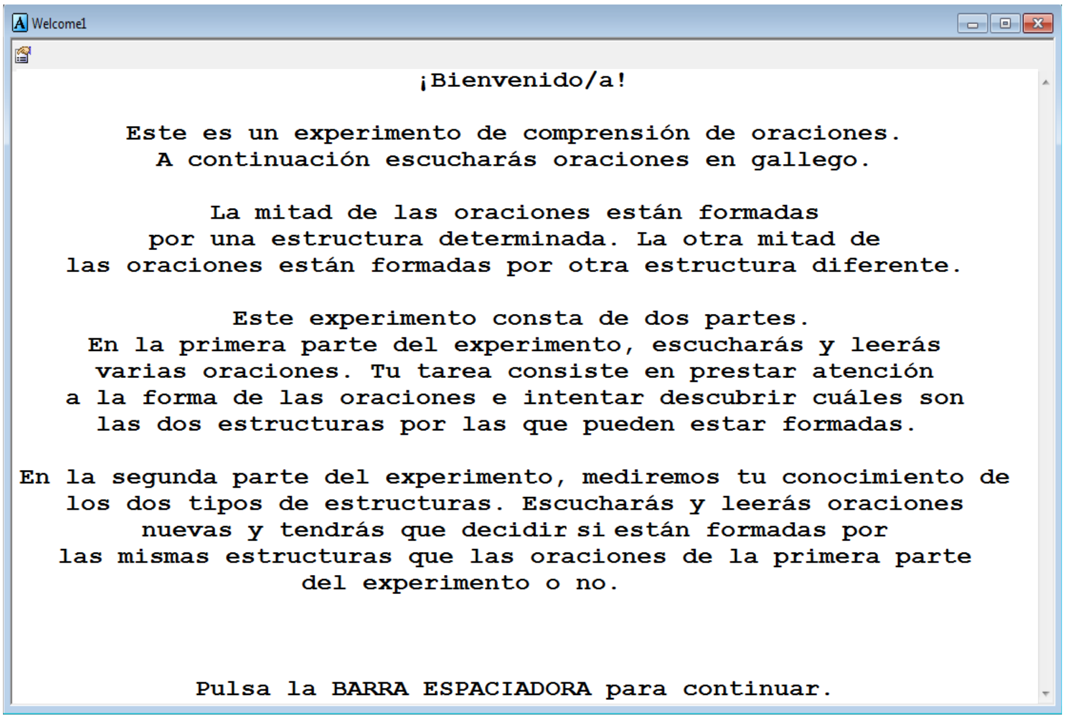
(79) a. É sorprendente que Ángela admire o lago.
b. É sorprendente Ángela admirar o lago.
c. *É sorprendente que Ángela admirar o lago.
d. *É sorprendente Ángela admire o lago.

(80) a. É necesario que Carlos fotografe a paisaxe.
b. É necesario Carlos fotografar a paisaxe.
c. *É necesario que Carlos fotografar a paisaxe.
d. *É necesario Carlos fotografe a paisaxe.

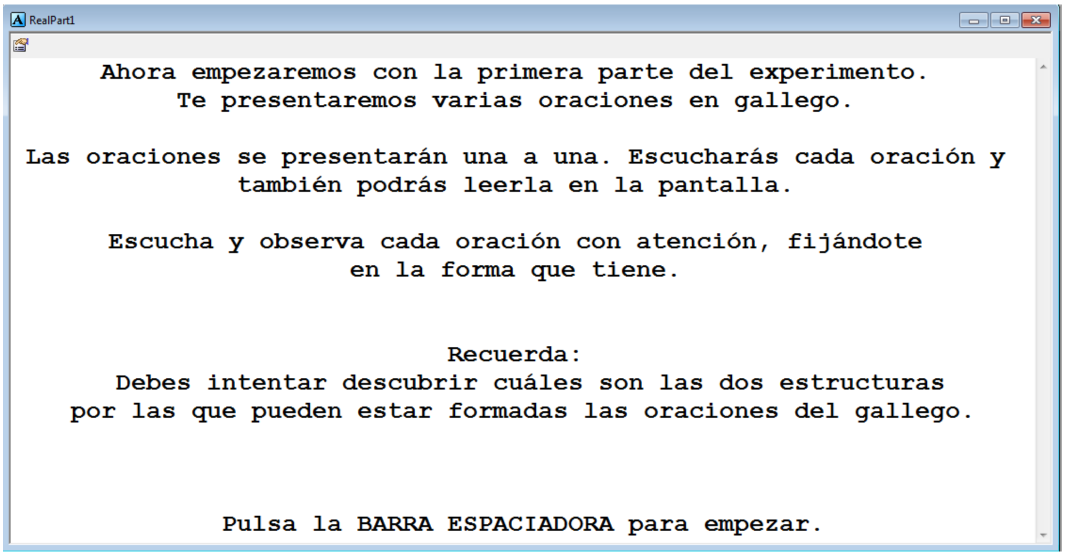
10. Instructions used in Experiment 2

This appendix includes the instructions for the exposure phase, the testing phase and the debriefing phase in Experiment 2. Instructions are presented in Spanish, the language of the experiment. The English translation is presented below each slide.

Exposure phase

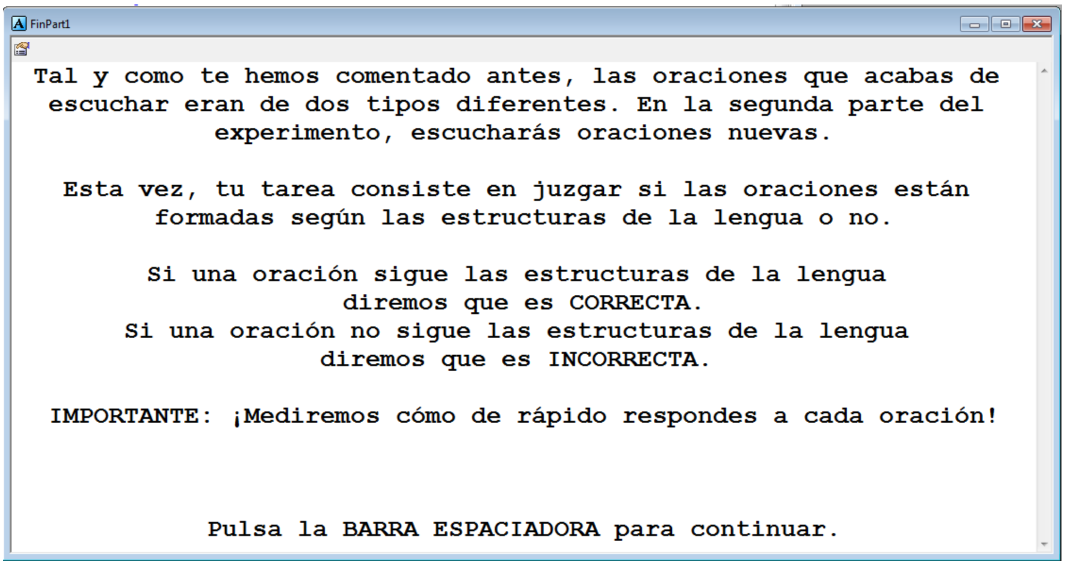


ENG: Welcome! This is an experiment on sentence comprehension. You will listen to some sentences in Galician. Half of the sentences are formed by a given structure. The other half are formed by a different structure. This experiment has two parts. In the first part, you will listen to and read several sentences. Your task is to pay attention to the form of the sentences and try to figure out which are the two structures by which they can be formed. In the second part of the experiment, you will be tested on your knowledge of the two structures. You will listen to and read new sentences and you will have to decide whether they are formed by the same structures as the sentences in the first part of the experiment or not. Press the SPACE BAR to continue.

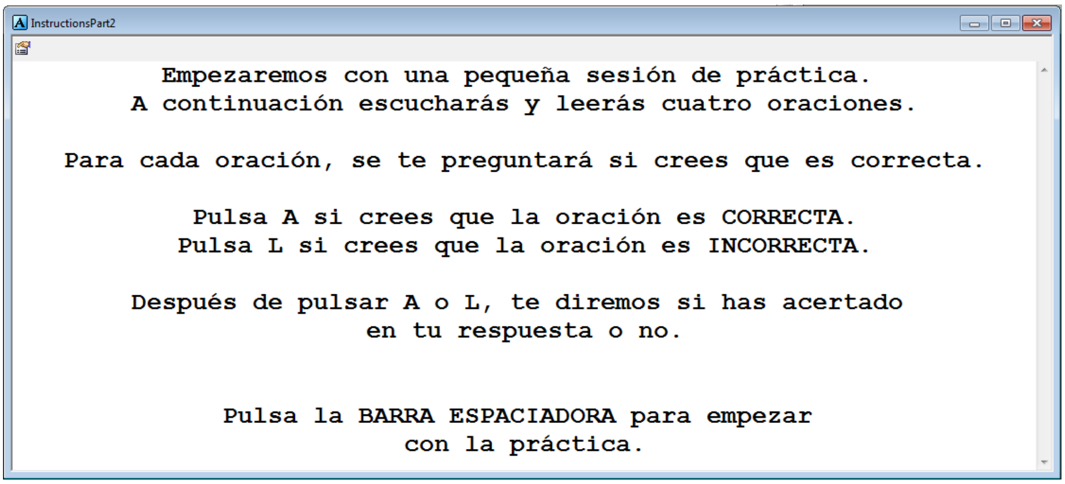


ENG: You will now start with the first part of the experiment. You will be presented with some sentences in Galician. The sentences will be presented one at a time. You will listen to each sentence and you will also be able to read it on the screen. Listen to and read each sentence carefully, paying attention to its form. Remember: You must try to figure out which are the two structures by which Galician sentences can be formed. Press the SPACE BAR to start.

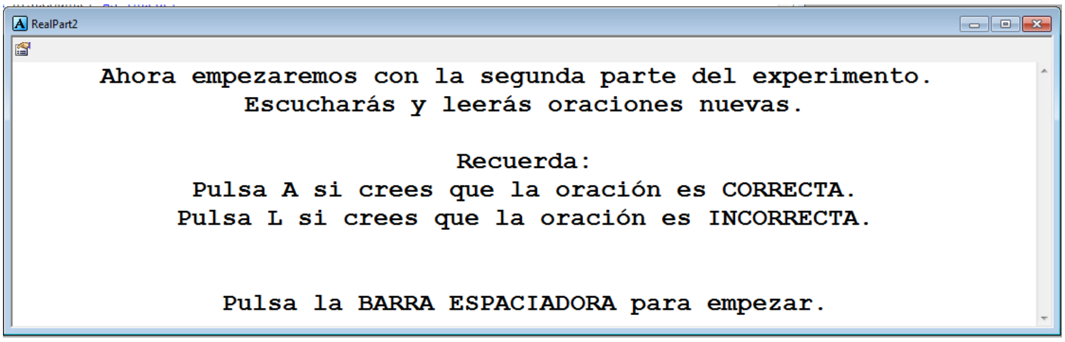
Testing phase



ENG: As mentioned, you listened to two different types of sentences. In the second part of the experiment, you will listen to new sentences. This time, your task is to judge whether the sentences are formed according to the structures of the language or not. If a sentence is formed according to one of the structures of the language, we say it is CORRECT. If a sentence is not formed by one of the structures of the language, we say it is INCORRECT. IMPORTANT: We will measure your response times! Press the SPACE BAR to start.

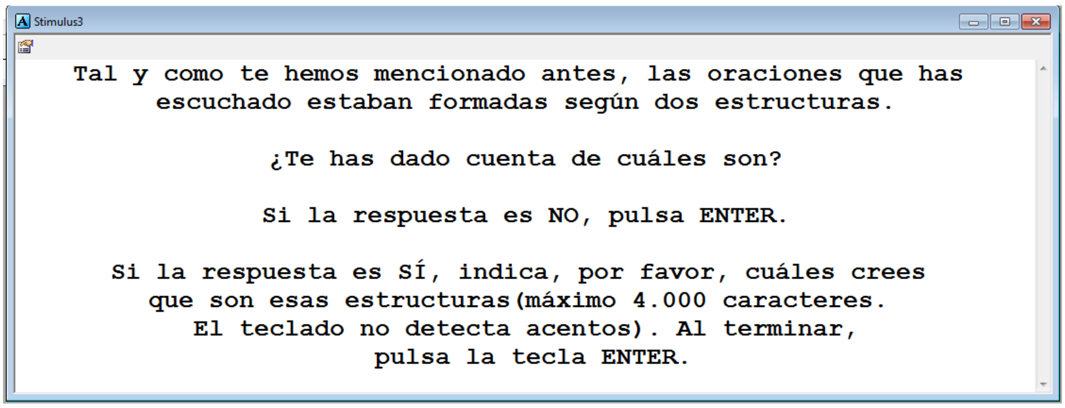


ENG: You will begin with a short practice session. You will listen to and read four sentences. For each sentence, you will be asked whether it is correct. Press A if you think that the sentence is CORRECT. Press L if you think that the sentence is INCORRECT. After pressing A or L, you will be told whether your answer was right. Press the SPACE BAR to begin with the practice.

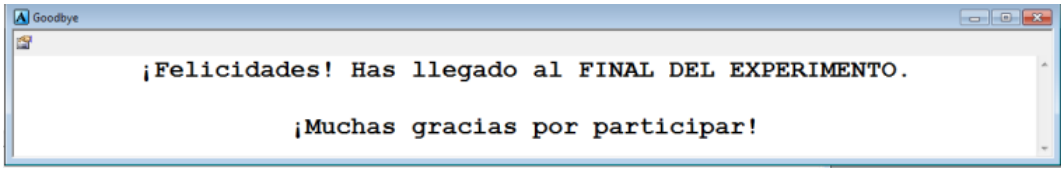


ENG: You will now start with the second part of the experiment. You will listen to and read new sentences. Remember: Press A if you think that the sentence is CORRECT. Press L if you think that the sentence is INCORRECT. Press the SPACE BAR to begin.

<u>Debriefing phase</u>



ENG: As previously mentioned, the sentences you listened to were formed according to two structures. Did you notice which structures were these? If your answer is NO, press ENTER. If your answer is YES, indicate, please, which you think these structures are (maximum 4,000 characters. The keyboard does not admit accent marks). When you finish, press ENTER.



ENG: Congratulations! You have reached the END OF THE EXPERIMENT. Thank you for participating!

11. Performance of learners who learnt the structures in Experiment 2

When just the subset of learners who evidenced learning of at least one of the two target structures was taken into account, the mean accuracy percentage for the grammatical and the ungrammatical similar structure rose to 93.50% (*SD* = 24.67%, 95%CI = [92.14, 94.85]). Similarly, mean accuracy for the grammatical and the ungrammatical dissimilar structure increased to 89.32% (*SD* = 30.89%, 95%CI = [87.63, 91.02]). Yet, participants continued to be significantly more accurate when judging the first pair of structures compared to the second (β = -0.68, *SE* = 0.25, *z* = -2.72, *p* < .01). The estimated coefficient of the effect of cross-linguistic similarity on accuracy, in log odds, was -0.68. This corresponds to an odds ratio of 0.50 to 1. Thus, the odds of judging a sentence correctly as opposed to incorrectly were 0.50 times smaller when the item was DS or *DS compared to when it was SS or *SS. In a similar vein, the mean *d'* scores reflecting sensitivity to SS vs *SS items and DS vs. *DS items increased to 3.17 (*SD* = 0.99, 95%CI = [3.12, 3.23]) and 2.78 (*SD* = 1.21, 95%CI = [2.71, 2.84]), respectively. Sensitivity to the former continued to be significantly greater than to the latter, as indicated by a paired-samples t-test, *t* (31) = 2.51, *p* = .02, *d* = 0.44[1]. In short, these results suggest that, also for this subset of learners, the similar structure was more firmly established in the linguistic system than the dissimilar one, in accordance with Hypothesis 2 (H2) proposed for Chapter 2's main research question.

Table A-11.1 shows mean accuracy percentages and *d'* scores for the grammatical and ungrammatical similar and dissimilar structures in the first and the last 20 test trials. As reported for the whole group of participants, accuracy and *d'* scores for SS-*SS items and DS-*DS items were significantly above chance in the first block of trials (all *p* <.001), right after the exposure phase. This denotes that, by that time, the two structures were part of learners' grammar and, thus, that the dissimilar structure was learnt during exposure, in agreement with Hypothesis 1 postulated for MQ1 (MQ1_H1). Next, I examined the effect of feedback on learning of the two structures. To this aim, a generalized linear mixed effects model and a 2x2 within subjects repeated-measures ANOVA tested for the interaction between the effect of Cross-linguistic similarity and Test block on accuracy and *d'* scores, respectively. The mixed model yielded an effect of Cross-linguistic similarity (β = 0.90, *SE* = 0.31, *z* = 2.96, *p* = .003), an effect of Test block (β = -1.75, *SE* = 0.45, *z* = -3.89, *p* <.001) and no interaction between the two variables (β = -0.39, *SE* = 0.54, *z* = -0.72, *p* = .47). The ANOVA obtained similar results, i.e. a main effect of Cross-linguistic similarity (*F* (1, 31) = 7.96, *p* = .008, large effect size of η_p^2 = 0.204), a main effect of Test block (*F* (1, 31) = 10.22, *p* = .003, large effect size of η_p^2 = 0.248) and no significant interaction (*F* (1, 31) = 0.70, *p* = .41, η_p^2 = 0.022). These results converge with those reported for the whole group of participants and receive the same interpretation, in line with Hypothesis 1 proposed for MQ2 (MQ2_H1).

[1] Non-parametric Wilcoxon signed-rank test: *V* = 292.5, *p* = .01, medium effect size of *r* = 0.45.

		Accuracy			*d'* scores		
		M	*SD (%)*	*95%CI*	*M*	*SD*	*95%CI*
First 20 test trials	SS - *SS	88.96	31.39	[85.49, 92.43]	2.09	1.05	[1.97, 2.20]
	DS - *DS	81.70	38.72	[77.42, 85.98]	1.63	1.16	[1.51, 1.76]
Last 20 test trials	SS - *SS	97.19	16.56	[95.37, 99.01]	2.51	0.35	[2.47, 2.55]
	DS - *DS	92.77	25.94	[89.90, 95.63]	2.28	0.92	[2.17, 2.38]

TABLE A-11.1. First and last 20 trials of the GJT in Experiment 2: Mean accuracy (%), *d'* scores, standard deviations and 95% confidence intervals for grammatical vs. ungrammatical similar and dissimilar structures in the subset of participants learning one or the two target structures. *M* = Mean, *SD* = Standard Deviation, *95%CI* = 95% Confidence Interval.

Appendix B

Appendices to Chapter 3

1. Linguistic information about the participants in Experiments 2 vs. 3

The tables below report the information obtained in the linguistic background questionnaire that participants in Experiments 2 and 3 filled out before the experiment. I report first the information about learners' language use in different life periods and in different environments. Then, I report learners' self-assessed proficiency in Spanish.

Experiment 2 (n = 44) vs. Experiment 3 (n = 44)

		Experiment 2	**Experiment 3**	**Two-sample t-tests**
Childhood	Primary school	2.30 (1.65)	2.68 (1.61)	$t(86) = -1.11$, $p = .27$, $d = -0.24$
	Home	1.07 (0.33)	1.20 (0.46)	$t(86) = -1.59$, $p = .12$, $d = -0.34$
	Other places	1.18 (0.39)	1.39 (0.75)	$t(86) = -1.60$, $p = .11$, $d = -0.34$
	Mean	1.52 (1.14)	1.76 (1.24)	$t(262) = -1.65$, $p = .10$, $d = -0.20$
Puberty	High school	2.36 (1.69)	2.75 (1.59)	$t(86) = -1.11$, $p = .27$, $d = -0.24$
	Home	1.07 (0.25)	1.18 (0.45)	$t(86) = -1.47$, $p = .15$, $d = -0.31$
	Other places	1.18 (0.39)	1.36 (0.65)	$t(86) = -1.59$, $p = .12$, $d = -0.34$
	Mean	1.54 (1.16)	1.77 (1.23)	$t(262) = -1.54$, $p = .12$, $d = -0.19$
Adulthood	University/work	1.34 (0.64)	1.39 (0.54)	$t(86) = -0.36$, $p = .72$, $d = -0.08$
	Home	1.05 (0.21)	1.16 (0.43)	$t(86) = -1.58$, $p = .12$, $d = -0.34$
	Other places	1.34 (0.61)	1.41 (0.62)	$t(86) = -0.52$, $p = .60$, $d = -0.11$
	Mean	1.24 (0.54)	1.32 (0.54)	$t(262) = -1.14$, $p = .26$, $d = -0.14$

TABLE B-1.1. Language use (SD in brackets) during childhood, puberty and adulthood in different environments as self-assessed by participants in Experiments 2 and 3. Scores are on a 7-point scale: 1 = Spanish only; 2 = Mostly Spanish, rarely Basque; 3 = Mostly Spanish, but Basque at least 25% of the time; 4 = Spanish and Basque with equal frequency; 5 = Mostly Basque, but Spanish at least 25% of the time; 6 = Mostly Basque, rarely Spanish; 7 = Basque only. Scores in Experiment 2 and Experiment 3 are compared by independent-samples t-tests. Cohen's d is reported as a standardised measure of effect size.

	Experiment 2	**Experiment 3**	**Two-sample t-tests**
Speaking	6.73 (0.59)	6.77 (0.42)	$t(86) = -0.41$, $p = .68$, $d = -0.09$
Listening	6.84 (0.37)	6.86 (0.35)	$t(86) = -0.30$, $p = .77$, $d = -0.06$
Reading	6.82 (0.50)	6.84 (0.37)	$t(86) = -0.24$, $p = .81$, $d = -0.05$
Writing	6.70 (0.55)	6.77 (0.42)	$t(86) = -0.65$, $p = .52$, $d = -0.14$
Mean	6.77 (0.51)	6.81 (0.39)	$t(350) = -0.82$, $p = .41$, $d = -0.09$

TABLE B-1.2. Proficiency level speaking, listening, reading and writing in Spanish (SD in brackets) as self-assessed by participants in Experiments 2 and 3. Scores are on a 7-point scale: 1 = Very poor; 2 = Poor; 3 = Enough; 4 = Pretty good; 5 = Good; 6 = Very good; 7 = Perfect. Scores in Experiment 2 and Experiment 3 are compared by independent-samples t-tests. Cohen's d is reported as a standardised measure of effect size.

2. Informed consent used in Experiment 3

This is the Spanish "Informed consent in comprehension tests" participants read and signed before Experiment 3. This consent provided participants with all the necessary information about the experiment so that they could decide freely and voluntarily whether they wanted to participate. The information given included: the project the study was part of, details of the Principal Investigators of the project and the person in charge of the experiment, description, aims and procedure of the study, risks and rights of the participant and policy of conservation and processing of personal data.

CONSENTIMIENTO INFORMADO EN PRUEBAS DE COMPRENSIÓN

El presente informe tiene como objetivo primordial proporcionarle toda la información necesaria sobre el experimento en el que va a participar y sobre la conservación y tratamiento de sus datos personales, con el objetivo de que pueda decidir libre y voluntariamente sobre su participación en el mismo.

Identificación del proyecto

Título del proyecto: Cross-linguistic influence in language learning, processing and aging
Financiación: Ministerio de Ciencia e Innovación
Título del estudio: The effect of frequency and syntactic co-activation in L2 syntax learning
Código del proyecto: PID2021-124056NB-I00

Identificación del investigador principal y forma de contacto

Nombre y apellidos: Mikel Santesteban y Kepa Erdozia
Dirección: Centro de Investigación Micaela Portilla 3.2. Dept. Lingüística y Estudios Vascos. Facultad de Letras, Universidad del País Vasco (UPV/EHU)
E-mail: mikel.santesteban@ehu.eus y kepa.erdozia@ehu.eus
Teléfono: 945013650

Identificación del investigador responsable

El investigador responsable se encargará de pasar la prueba experimental y de informarle adecuadamente.
Nombre y apellidos: Noèlia Sanahuja Cobacho
Dirección: Centro de Investigación Micaela Portilla 3.2. Dept. Lingüística y Estudios Vascos (UPV/EHU)
E-mail: noelia.sanahuja@ehu.eus

DESCRIPCIÓN Y OBJETIVOS DE LA INVESTIGACIÓN

Nuestro objetivo principal es llegar a entender cómo la activación interlingüística de las dos lenguas modula el aprendizaje y el procesamiento de la L2 en el hablante bilingüe. Dentro de este objetivo general trabajaremos con la hipótesis de que la activación interlingüística guía la mayoría de los procesos cuando utilizamos una segunda lengua. Investigaremos cual es el papel de la distancia tipológica entre L1 y L2 en los efectos de interferencia interlingüística durante el aprendizaje y procesamiento de lenguaje y el papel del conocimiento sintáctico en el aprendizaje de reglas de L2.

PROPÓSITO DEL ESTUDIO

El objetivo del presente estudio es investigar cómo el cerebro procesa el lenguaje durante el aprendizaje de una segunda lengua.

PROCEDIMIENTO

El procedimiento a seguir consiste en escuchar y leer unas oraciones mientras se mira la pantalla de un ordenador. Seguidamente, se realizará una prueba lingüística basada en las oraciones que se han escuchado. El experimento se realizará en 1 sesión y tendrá una duración máxima de 75 minutos.

Riesgos e incomodidades

Ninguno de los procedimientos representa peligro alguno para la salud o integridad física. Todas las intervenciones se llevarán a cabo con todas las medidas preventivas requeridas en la situación de Covid-19.

DERECHOS DEL PARTICIPANTE

Cláusula de voluntariedad y derecho de revocación

La información que contienen los datos personales del participante o cualquier otro dato identificativo no se proporcionará a terceros y se protegerá la privacidad de los mismos. Los resultados de este proyecto pueden llegar a publicarse en libros o revistas especializadas o pueden usarse con finalidades didácticas. La participación en este estudio es completamente voluntaria y, como tal, puede revocar el consentimiento dado en cualquier momento, sin dar explicaciones de ningún tipo y sin que ello suponga ningún perjuicio o medida en su contra. De igual forma, a criterio del investigador, usted puede ser retirado del estudio por alguna de las siguientes razones: (a) si no cumple con los requisitos mínimos que se establezcan para participar en el estudio; (b) si por cualquier motivo se interrumpe el estudio.

Cláusula sobre el derecho a tener más información sobre el proyecto

Si colabora en este estudio, una vez haya finalizado, tendrá usted a su disposición toda la información relativa a los resultados obtenidos en el mismo. Para acceder a ella, es necesario que se ponga en contacto con el investigador responsable del proyecto a través de la dirección de e-mail que consta en este documento.

PROTECCIÓN DE DATOS:

Se le informa de que de conformidad al Reglamento Europeo de Protección de Datos (UE2016/679):

- Los datos personales que se le solicitan son:
  - a) Datos de carácter identificativo: DNI/NIF, NOMBRE Y APELLIDOS, DIRECCIÓN (POSTAL, ELECTRÓNICA), TELÉFONO, IMAGEN/VOZ
  - b) Datos de características personales: FECHA DE NACIMIENTO, LUGAR DE NACIMIENTO, EDAD, SEXO, NACIONALIDAD, LENGUA MATERNA
  - c) Datos académicos y profesionales: FORMACIÓN, TITULACIONES
- El código del tratamiento de datos es: TI0091
- El nombre del tratamiento de datos es: DATOS GOGO ELEBIDUNA-MENTE BILINGÜE
- La finalidad de este tratamiento es: CUESTIONARIO DE PERFIL LINGÜÍSTICO DE LOS PARTICIPANTES EN LOS EXPERIMENTOS PSICOLINGÜÍSTICOS DEL GRUPO DE INVESTIGACIÓN "GOGO ELEBIDUNA/MENTE BILINGÜE"
- El responsable del tratamiento de datos es la UPV/EHU:

Identidad: Universidad del País Vasco/Euskal Herriko Unibertsitatea CIF: Q4818001B Dirección postal: Barrio Sarriena s/n, 48940-Leioa (Bizkaia) Página web: www.ehu.eus Datos de contacto del Delegado de Protección de Datos: dpd@ehu.eus

- El periodo de conservación de sus datos será: Los datos se conservarán mientras no se solicite su supresión por la persona interesada y, en cualquier caso, siempre que estén abiertos los plazos de recurso y/o reclamación procedente o mientras sigan respondiendo a la finalidad para la que fueron obtenidos.
- La legitimación del tratamiento es: su consentimiento informado.
- Cesiones y transferencias internacionales de sus datos: No se cederán datos salvo previsión legal. No se efectuarán transferencias internacionales.
- Los derechos sobre sus datos son los de acceso, supresión, rectificación, oposición, limitación del tratamiento, portabilidad y olvido. Puede ejercerlos enviando su petición a dpd@ehu.eus.
- Tiene a su disposición información adicional en http://www.ehu.eus/babestu
- La información completa sobre este tratamiento está en: https://www.ehu.eus/es/web/idazkaritza-nagusia/ikerketa-datu-pertsonalen-tratamenduak

IDENTIFICACIÓN DE LA PERSONA QUE PRESTA EL CONSENTIMIENTO

Yo, ____, con DNI nº ___________________ declaro que he leído este documento y que doy mi consentimiento a participar voluntariamente en este estudio.

____ ___________________

Voluntario/a Fecha

El investigador abajo firmante declara que el participante ha recibido la información escrita y oral necesaria para garantizar que su participación pueda considerarse libre y voluntaria.

____ ___________________

Investigador/a Fecha

3. List of verbs used in Experiment 3

Exposure set	Testing set
Desmontar	Transcribir
Sp. “desmontar” /Eng. “disassemble”	Sp. “transcribir” /Eng. “transcribe”
Rifar	Emendar
Sp. “rifar” /Eng. “raffle off”	Sp. “enmendar” /Eng. “amend”
Silenciar	Rememorar
Sp. “silenciar” /Eng. “mute”	Sp. “rememorar” /Eng. “reminisce about”
Reciclar	Simplificar
Sp. “reciclar” /Eng. “recycle”	Sp. “simplificar” /Eng. “simplify”
Enchufar	Escanear
Sp. “enchufar” /Eng. “plug in”	Sp. “escanear” /Eng. “scan”
Degustar	Tramitar
Sp. “degustar” /Eng. “taste”	Sp. “tramitar” /Eng. “process”
Engulir	Abreviar
Sp. “engullir” /Eng. “gobble up”	Sp. “abreviar” /Eng. “abbreviate”
Humedecer	Rectificar
Sp. “humedecer” /Eng. “moisten”	Sp. “rectificar” /Eng. “rectify”
Recubrir	Validar
Sp. “recubrir” /Eng. “coat”	Sp. “validar” /Eng. “validate”
Destapar	Falsear
Sp. “destapar” /Eng. “uncover”	Sp. “falsear” /Eng. “falsify”
Pospoñer	Ventilar
Sp. “posponer” /Eng. “postpone”	Sp. “ventilar” /Eng. “ventilate”
Arbitrar	Desaloxar
Sp. “arbitrar” /Eng. “referee”	Sp. “desalojar” /Eng. “vacate”
Custear	Fregar
Sp. “costear” /Eng. “fund, pay for”	Sp. “fregar” /Eng. “wash, mop”
Invalidar	Exaltar
Sp. “invalidar” /Eng. “invalidate”	Sp. “exaltar” /Eng. “exalt”
Atrasar	Amoblar
Sp. “atrasar” /Eng. “delay”	Sp. “amueblar” /Eng. “furnish”
Embalar	Eloxiar
Sp. “embalar” /Eng. “pack”	Sp. “elogiar” /Eng. “praise”
Obsequiar	Desatender
Sp. “obsequiar” /Eng. “give away”	Sp. “desatender” /Eng. “neglect”
Ofertar	Custodiar
Sp. “ofertar” /Eng. “offer”	Sp. “custodiar” /Eng. “guard”
Abaratar	Embelecer
Sp. “abaratar” /Eng. “cheapen”	Sp. “embellecer” /Eng. “beautify”
Acortar	Arrendar
Sp. “acortar” /Eng. “shorten”	Sp. “arrendar” /Eng. “lease”

TABLE B-3.1. Verbs in the exposure set and the testing set of Experiment 3. Spanish (Sp.) and English (Eng.) translations are provided below each word.

4. Lexical characteristics of the exposure set across conditions in Experiment 3

Measure	**SS**	**DS**	**Two-sample t-tests**
Length proper nouns	5.3 (0.84)	5.2 (0.83)	$t(98) = 0.60$, $p = .55$
Length verbs	7.8 (1.03)	7.8 (0.95)	$t(98) = 0$, $p = 1$
Length inanimate nouns	6.54 (1.75)	6.26 (1.70)	$t(98) = 0.81$, $p = .42$
Phonological LD proper nouns	0.26 (0.52)	0.24 (0.55)	$t(98) = 0.18$, p = .85
Orthographic LD proper nouns	0	0	X
Phonological LD verbs	0.24 (0.43)	0.26 (0.44)	$t(98) = -0.23$, $p = .82$
Orthographic LD verbs	0.16 (0.37)	0.14 (0.56)	$t(98) = 0.28$, $p = .78$
Phonological LD inanimate nouns	1.28 (0.78)	1.32 (0.79)	$t(98) = -0.25$, $p = .80$
Orthographic LD inanimate nouns	0.78 (1.02)	0.92 (1.12)	$t(98) = -0.65$, $p = .51$
Frequency proper nouns (CORPES XXI)	18.16 (9.66)	18.65 (9.65)	$t(98) = -0.25$, $p = .80$
Frequency verbs (CORPES XXI)	4.67 (2.01)	4.85 (1.97)	$t(98) = -0.45$, $p = .65$
Frequency inanimate nouns (CORPES XXI)	57.51 (62.22)	47.02 (55.60)	$t(98) = 0.89$, $p = .38$

TABLE B-4.1. Mean length (number of letters), phonological and orthographic overlap with Spanish translations (Levenshtein distance, LD) and frequency per million in Spanish in the CORPES XXI (SD in brackets) for proper nouns, verbs and inanimate nouns in sentences formed by the similar structure (SS) and the dissimilar structure (DS) in Experiment 3's two exposure lists. Independent-samples t-tests compare each variable in the two conditions. The descriptive data and t-tests are valid for the two exposure lists, since the verbs and nouns in the SS condition in list 1 were the verbs and nouns in the DS condition in list 2, and vice versa.

5. Experimental materials used in Experiment 3

Exposure set

Galician-based sentences constituting the exposure set in Experiment 3. Sentences are presented in pairs, corresponding to the two conditions in the exposure phase: a. (Grammatical) similar structure and b. (Grammatical) dissimilar structure. I provide the English translation shared for each pair of sentences.

(1) a. É importante que Pedro desmonte a radio.
b. É importante Pedro desmontar a radio.
"It is important that Pedro disassembles the radio."

(2) a. É posible que Mónica rife a consola.
b. É posible Mónica rifar a consola.
"It is possible that Mónica raffles off the console."

(3) a. É probable que Ángel silencie o móbil.
b. É probable Ángel silenciar o móbil.
"It is probable that Ángel mutes the mobile."

(4) a. É sorprendente que Isabel recicle o portátil.
b. É sorprendente Isabel reciclar o portátil.
"It is surprising that Isabel recycles the laptop."

(5) a. É necesario que Pablo enchufe o televisor.
b. É necesario Pablo enchufar o televisor.
"It is necessary that Pablo plugs in the television."

(6) a. É importante que Lucía deguste a torta.
b. É importante Lucía degustar a torta.
"It is important that Lucía tastes the cake."

(7) a. É posible que Luis engula a carne.
b. É posible Luis engulir a carne.
"It is possible that Luis gobbles up the meat."

(8) a. É probable que Paula humedeza o pan.
b. É probable Paula humedecer o pan.
"It is probable that Paula moistens the bread."

(9) a. É sorprendente que Jorge recubra o brócoli.
b. É sorprendente Jorge recubrir o brócoli.
"It is surprising that Jorge coats the broccoli."

(10) a. É necesario que Alicia destape a cenoria.
b. É necesario Alicia destapar a cenoria.
"It is necessary that Alicia uncovers the carrot."

(11) a. É importante que Alberto pospoña o concurso.
b. É importante Alberto pospoñer o concurso.
"It is important that Alberto postpones the contest."

(12) a. É posible que Silvia arbitre a proba.
b. É posible Silvia arbitrar a proba.
"It is possible that Silvia referees the test."

(13) a. É probable que Adrián custee o torneo.
b. É probable Adrián custear o torneo.
"It is probable that Adrián funds the tournament."

(14) a. É sorprendente que Irene invalide a carreira.
b. É sorprendente Irene invalidar a carreira.
"It is surprising that Irene invalidates the race."

(15) a. É necesario que Iván atrase o campionato.
b. É necesario Iván atrasar o campionato.
"It is necessary that Iván delays the championship."

(16) a. É importante que Rosa embale a xoia.
b. É importante Rosa embalar a xoia.
"It is important that Rosa packs the jewel."

(17) a. É posible que Raúl obsequie a pulseira.
b. É posible Raúl obsequiar a pulseira.
"It is possible that Raúl gives away the bracelet."

(18) a. É probable que Andrea oferte o abrigo.
b. É probable Andrea ofertar o abrigo.
"It is probable that Andrea offers the coat."

(19) a. É sorprendente que Óscar abarate o reloxo.
b. É sorprendente Óscar abaratar o reloxo.
"It is surprising that Óscar cheapens the watch."

(20) a. É necesario que Carmen acorte o xersei.
b. É necesario Carmen acortar o xersei.
"It is necessary that Carmen shortens the sweater."

(21) a. É importante que Pablo recicle o móbil.
b. É importante Pablo reciclar o móbil.
"It is important that Pablo recycles the mobile."

(22) a. É posible que Pedro enchufe o portátil.
b. É posible Pedro enchufar o portátil.
"It is possible that Pedro plugs in the laptop."

(23) a. É probable que Mónica desmonte o televisor.
b. É probable Mónica desmontar o televisor.
"It is probable that Mónica disassembles the television."

(24) a. É sorprendente que Ángel rife a radio.
b. É sorprendente Ángel rifar a radio.
"It is surprising that Ángel raffles off the radio."

(25) a. É necesario que Isabel silencie a consola.
b. É necesario Isabel silenciar a consola.
"It is necessary that Isabel mutes the console."

(26) a. É importante que Alicia recubra o pan.
b. É importante Alicia recubrir o pan.
"It is important that Alicia coats the bread."

(27) a. É posible que Lucía destape o brócoli.
b. É posible Lucía destapar o brócoli.
"It is possible that Lucía uncovers the broccoli."

(28) a. É probable que Luis deguste a cenoria.
b. É probable Luis degustar a cenoria.
"It is probable that Luis tastes the carrot."

(29) a. É sorprendente que Paula engula a torta.
b. É sorprendente Paula engulir a torta.
"It is surprising that Paula gobbles up the cake."

(30) a. É necesario que Jorge humedeza a carne.
b. É necesario Jorge humedecer a carne.
"It is necessary that Jorge moistens the meat."

(31) a. É importante que Iván invalide o torneo.
b. É importante Iván invalidar o torneo.
"It is important that Iván invalidates the tournament."

(32) a. É posible que Alberto atrase a carreira.
b. É posible Alberto atrasar a carreira.
"It is possible that Alberto delays the race."

(33) a. É probable que Silvia pospoña o campionato.
b. É probable Silvia pospoñer o campionato.
"It is probable that Silvia postpones the championship."

(34) a. É sorprendente que Adrián arbitre o concurso.
b. É sorprendente Adrián arbitrar o concurso.
"It is surprising that Adrián referees the contest."

(35) a. É necesario que Irene custee a proba.
b. É necesario Irene custear a proba.
"It is necessary that Irene pays for the test."

(36) a. É importante que Carmen abarate o abrigo.
b. É importante Carmen abaratar o abrigo.
"It is important that Carmen cheapens the coat."

(37) a. É posible que Rosa acorte o reloxo.
b. É posible Rosa acortar o reloxo.
"It is possible that Rosa shortens the watch."

(38) a. É probable que Raúl embale o xersei.
b. É probable Raúl embalar o xersei.
"It is probable that Raúl packs the sweater."

(39) a. É sorprendente que Andrea obsequie a xoia.
b. É sorprendente Andrea obsequiar a xoia.
"It is surprising that Andrea gives away the jewel."

(40) a. É necesario que Óscar oferte a pulseira.
b. É necesario Óscar ofertar a pulseira.
"It is necessary that Óscar offers the bracelet."

(41) a. É importante que Isabel rife o televisor.
b. É importante Isabel rifar o televisor.
"It is important that Isabel raffles off the television."

(42) a. É posible que Pablo silencie a radio.
b. É posible Pablo silenciar a radio.
"It is possible that Pablo mutes the radio."

(43) a. É probable que Pedro recicle a consola.
b. É probable Pedro reciclar a consola.
"It is probable that Pedro recycles the console."

(44) a. É sorprendente que Mónica enchufe o móbil.
b. É sorprendente Mónica enchufar o móbil.
"It is surprising that Mónica plugs in the mobile."

(45) a. É necesario que Ángel desmonte o portátil.
b. É necesario Ángel desmontar o portátil.
"It is necessary that Ángel disassembles the laptop."

(46) a. É importante que Jorge engula a cenoria.
b. É importante Jorge engulir a cenoria.
"It is important that Jorge gobbles up the carrot."

(47) a. É posible que Alicia humedeza a torta.
b. É posible Alicia humedecer a torta.
"It is possible that Alicia moistens the cake."

(48) a. É probable que Lucía recubra a carne.
b. É probable Lucía recubrir a carne.
"It is probable that Lucía coats the meat."

(49) a. É sorprendente que Luis destape o pan.
b. É sorprendente Luis destapar o pan.
"It is surprising that Luis uncovers the bread."

(50) a. É necesario que Paula deguste o brócoli.
b. É necesario Paula degustar o brócoli.
"It is necessary that Paula tastes the broccoli."

(51) a. É importante que Irene arbitre o campionato.
b. É importante Irene arbitrar o campionato.
"It is important that Irene referees the championship."

(52) a. É posible que Iván custee o concurso.
b. É posible Iván custear o concurso.
"It is possible that Iván funds the contest."

(53) a. É probable que Alberto invalide a proba.
b. É probable Alberto invalidar a proba.
"It is probable that Alberto invalidates the test."

(54) a. É sorprendente que Silvia atrase o torneo.
b. É sorprendente Silvia atrasar o torneo.
"It is surprising that Silvia delays the tournament."

(55) a. É necesario que Adrián pospoña a carreira.
b. É necesario Adrián pospoñer a carreira.
"It is necessary that Adrián postpones the race."

(56) a. É importante que Óscar obsequie o xersei.
b. É importante Óscar obsequiar o xersei.
"It is important that Óscar gives away the sweater."

(57) a. É posible que Carmen oferte a xoia.
b. É posible Carmen ofertar a xoia.
"It is possible that Carmen offers the jewel."

(58) a. É probable que Rosa abarate a pulseira.
b. É probable Rosa abaratar a pulseira.
"It is probable that Rosa cheapens the bracelet."

(59) a. É sorprendente que Raúl acorte o abrigo.
b. É sorprendente Raúl acortar o abrigo.
"It is surprising that Raúl shortens the coat."

(60) a. É necesario que Andrea embale o reloxo.
b. É necesario Andrea embalar o reloxo.
"It is necessary that Andrea packs the watch."

(61) a. É importante que Ángel enchufe a consola.
b. É importante Ángel enchufar a consola.
"It is important that Ángel plugs in the console."

(62) a. É posible que Isabel desmonte o móbil.
b. É posible Isabel desmontar o móbil.
"It is possible that Isabel disassembles the mobile."

(63) a. É probable que Pablo rife o portátil.
b. É probable Pablo rifar o portátil.
"It is probable that Pablo raffles off the laptop."

(64) a. É sorprendente que Pedro silencie o televisor.
b. É sorprendente Pedro silenciar o televisor.
"It is surprisng that Pedro mutes the television."

(65) a. É necesario que Mónica recicle a radio.
b. É necesario Mónica reciclar a radio.
"It is necessary that Mónica recycles the radio."

(66) a. É importante que Paula destape a carne.
b. É importante Paula destapar a carne.
"It is important that Paula uncovers the meat."

(67) a. É posible que Jorge deguste o pan.
b. É posible Jorge degustar o pan.
"It is possible that Jorge tastes the bread."

(68) a. É probable que Alicia engula o brócoli.
b. É probable Alicia engulir o brócoli.
"It is probable that Alicia gobbles up the broccoli."

(69) a. É sorprendente que Lucía humedeza a cenoria.
b. É sorprendente Lucía humedecer a cenoria.
"It is surprising that Lucía moistens the carrot."

(70) a. É necesario que Luis recubra a torta.
b. É necesario Luis recubrir a torta.
"It is necessary that Luis coats the cake."

(71) a. É importante que Adrián atrase a proba.
b. É importante Adrián atrasar a proba.
"It is important that Adrián delays the test."

(72) a. É posible que Irene pospoña o torneo.
b. É posible Irene pospoñer o torneo.
"It is possible that Irene postpones the tournament."

(73) a. É probable que Iván arbitre a carreira.
b. É probable Iván arbitrar a carreira.
"It is probable that Iván referees the race."

(74) a. É sorprendente que Alberto custee o campionato.
b. É sorprendente Alberto custear o campionato.
"It is surprising that Alberto funds the championship."

(75) a. É necesario que Silvia invalide o concurso.
b. É necesario Silvia invalidar o concurso.
"It is necessary that Silvia invalidates the contest."

(76) a. É importante que Andrea acorte a pulseira.
b. É importante Andrea acortar a pulseira.
"It is important that Andrea shortens the bracelet."

(77) a. É posible que Óscar embale o abrigo.
b. É posible Óscar embalar o abrigo.
"It is possible that Óscar packs the coat."

(78) a. É probable que Carmen obsequie o reloxo.
b. É probable Carmen obsequiar o reloxo.
"It is probable that Carmen gives away the watch."

(79) a. É sorprendente que Rosa oferte o xersei.
b. É sorprendente Rosa ofertar o xersei.
"It is surprising that Rosa offers the sweater."

(80) a. É necesario que Raúl abarate a xoia.
b. É necesario Raúl abaratar a xoia.
"It is necessary that Raúl cheapens the jewel."

(81) a. É importante que Mónica silencie o portátil.
b. É importante Mónica silenciar o portátil.
"It is important that Mónica mutes the laptop."

(82) a. É posible que Ángel recicle o televisor.
b. É posible Ángel reciclar o televisor.
"It is possible that Ángel recycles the television."

(83) a. É probable que Isabel enchufe a radio.
b. É probable Isabel enchufar a radio.
"It is probable that Isabel plugs in the radio."

(84) a. É sorprendente que Pablo desmonte a consola.
b. É sorprendente Pablo desmontar a consola.
"It is surprising that Pablo disassembles the console."

(85) a. É necesario que Pedro rife o móbil.
b. É necesario Pedro rifar o móbil.
"It is necessary that Pedro raffles off the mobile."

(86) a. É importante que Luis humedeza o brócoli.
b. É importante Luis humedecer o brócoli.
"It is important that Luis moistens the broccoli."

(87) a. É posible que Paula recubra a cenoria.
b. É posible Paula recubrir a cenoria.
"It is possible that Paula coats the carrot."

(88) a. É probable que Jorge destape a torta.
b. É probable Jorge destapar a torta.
"It is probable that Jorge uncovers the cake."

(89) a. É sorprendente que Alicia deguste a carne.
b. É sorprendente Alicia degustar a carne.
"It is surprising that Alicia tastes the meat."

(90) a. É necesario que Lucía engula o pan.
b. É necesario Lucía engulir o pan.
"It is necessary that Lucía gobbles up the bread."

(91) a. É importante que Silvia custee a carreira.
b. É importante Silvia custear a carreira.
"It is important that Silvia funds the race."

(92) a. É posible que Adrián invalide o campionato.
b. É posible Adrián invalidar o campionato.
"It is possible that Adrián invalidates the championship."

(93) a. É probable que Irene atrase o concurso.
b. É probable Irene atrasar o concurso.
"It is probable that Irene delays the contest."

(94) a. É sorprendente que Iván pospoña a proba.
b. É sorprendente Iván pospoñer a proba.
"It is surprising that Iván postpones the test."

(95) a. É necesario que Alberto arbitre o torneo.
b. É necesario Alberto arbitrar o torneo.
"It is necessary that Alberto referees the tournament."

(96) a. É importante que Raúl oferte o reloxo.
b. É importante Raúl ofertar o reloxo.
"It is important that Raúl offers the watch."

(97) a. É posible que Andrea abarate o xersei.
b. É posible Andrea abaratar o xersei.
"It is possible that Andrea cheapens the sweater."

(98) a. É probable que Óscar acorte a xoia.
b. É probable Óscar acortar a xoia.
"It is probable that Óscar shortens the jewel."

(99) a. É sorprendente que Carmen embale a pulseira.
b. É sorprendente Carmen embalar a pulseira.
"It is surprising that Carmen packs the bracelet."

(100) a. É necesario que Rosa obsequie o abrigo.
b. É necesario Rosa obsequiar o abrigo.
"It is necessary that Rosa gives away the coat."

<u>Testing set</u>

Galician-based sentences constituting the testing set in Experiment 3. Sentences are presented in groups of four, corresponding to the four conditions in the testing phase: a. Grammatical similar structure, b. Grammatical dissimilar structure, c. Ungrammatical similar structure and d. Ungrammatical dissimilar structure. I provide the English translation shared for each group of sentences (intended meaning for all sentences irrespective of their grammaticality).

(1) a. É importante que Antonio transcriba a carta.
b. É importante Antonio transcribir a carta.
c. *É importante que Antonio transcribir a carta.
d. *É importante Antonio transcriba a carta.
"It is important that Antonio transcribes the letter."

(2) a. É posible que Sofía emende o informe.
b. É posible Sofía emendar o informe.
c. *É posible que Sofía emendar o informe.
d. *É posible Sofía emende o informe.
”It is possible that Sofía amends the report."

(3) a. É probable que José rememore a postal.
b. É probable José rememorar a postal.
c. *É probable que José rememorar a postal.
d. *É probable José rememore a postal.
"It is probable that José reminisces about the postcard."

(4) a. É sorprendente que Ana simplifique o correo.
b. É sorprendente Ana simplificar o correo.
c. *É sorprendente que Ana simplificar o correo.
d. *É sorprendente Ana simplifique o correo.
"It is surprising that Ana simplifies the email."

(5) a. É necesario que Manuel escanee a novela.
b. É necesario Manuel escanear a novela.
c. *É necesario que Manuel escanear a novela.
d. *É necesario Manuel escanee a novela.
"It is necessary that Manuel scans the novel."

(6) a. É importante que Laura tramite a factura.
b. É importante Laura tramitar a factura.
c. *É importante que Laura tramitar a factura.
d. *É importante Laura tramite a factura.
"It is important that Laura processes the invoice."

(7) a. É posible que Francisco abrevie o libro.
b. É posible Francisco abreviar o libro.
c. *É posible que Francisco abreviar o libro.
d. *É posible Francisco abrevie o libro.
"It is possible that Francisco abbreviates the book."

(8) a. É probable que Victoria rectifique o comunicado.
b. É probable Victoria rectificar o comunicado.
c. *É probable que Victoria rectificar o comunicado.
d. *É probable Victoria rectifique o comunicado.
"It is probable that Victoria rectifies the statement."

(9) a. É sorprendente que Juan valide a proposta.
b. É sorprendente Juan validar a proposta.
c. *É sorprendente que Juan validar a proposta.
d. *É sorprendente Juan valide a proposta.
"It is surprising that Juan validates the proposal."

(10) a. É necesario que Marta falsee a lei.
b. É necesario Marta falsear a lei.
c. *É necesario que Marta falsear a lei.
d. *É necesario Marta falsee a lei.
"It is necessary that Marta falsifies the law."

(11) a. É importante que Alejandro ventile a casa.
b. É importante Alejandro ventilar a casa.
c. *É importante que Alejandro ventilar a casa.
d. *É importante Alejandro ventile a casa.
"It is important that Alejandro ventilates the house."

(12) a. É posible que Elena desaloxe a habitación.
b. É posible Elena desaloxar a habitación.
c. *É posible que Elena desaloxar a habitación.
d. *É posible Elena desaloxe a habitación.
"It is possible that Elena vacates the room."

(13) a. É probable que Javier fregue a cociña.
b. É probable Javier fregar a cociña.
c. *É probable que Javier fregar a cociña.
d. *É probable Javier fregue a cociña.
"It is probable that Javier mops the kitchen."

(14) a. É sorprendente que Sara exalte o lago.
b. É sorprendente Sara exaltar o lago.
c. *É sorprendente que Sara exaltar o lago.
d. *É sorprendente Sara exalte o lago.
"It is surprising that Sara exalts the lake."

(15) a. É necesario que Salvador amoble o apartamento.
b. É necesario Salvador amoblar o apartamento.
c. *É necesario que Salvador amoblar o apartamento.
d. *É necesario Salvador amoble o apartamento.
"It is necessary that Salvador furnishes the apartment."

(16) a. É importante que Ángela eloxie a paisaxe.
b. É importante Ángela eloxiar a paisaxe.
c. *É importante que Ángela eloxiar a paisaxe.
d. *É importante Ángela eloxie a paisaxe.
"It is important that Ángela praises the landscape."

(17) a. É posible que Carlos desatenda o cadro.
b. É posible Carlos desatender o cadro.
c. *É posible que Carlos desatender o cadro.
d. *É posible Carlos desatenda o cadro.
"It is possible that Carlos neglects the picture."

(18) a. É probable que Julia custodie a escultura.
b. É probable Julia custodiar a escultura.
c. *É probable que Julia custodiar a escultura.
d. *É probable Julia custodie a escultura.
"It is probable that Julia guards the sculpture."

(19) a. É sorprendente que Miguel embeleza o xardín.
b. É sorprendente Miguel embelecer o xardín.
c. *É sorprendente que Miguel embelecer o xardín.
d. *É sorprendente Miguel embeleza o xardín.
"It is surprising that Miguel beautifies the garden."

(20) a. É necesario que Alba arrende o almacén.
b. É necesario Alba arrendar o almacén.
c. *É necesario que Alba arrendar o almacén.
d. *É necesario Alba arrende o almacén.
"It is necessary that Alba leases the warehouse."

(21) a. É importante que Manuel simplifique a postal.
b. É importante Manuel simplificar a postal.
c. *É importante que Manuel simplificar a postal.
d. *É importante Manuel simplifique a postal.
"It is important that Manuel simplifies the postcard."

(22) a. É posible que Antonio escanee o correo.
b. É posible Antonio escanear o correo.
c. *É posible que Antonio escanear o correo.
d. *É posible Antonio escanee o correo.
"It is possible that Antonio scans the email."

(23) a. É probable que Sofía transcriba a novela.
b. É probable Sofía transcribir a novela.
c. *É probable que Sofía transcribir a novela.
d. *É probable Sofía transcriba a novela.
"It is probable that Sofía transcribes the novel."

(24) a. É sorprendente que José emende a factura.
b. É sorprendente José emendar a factura.
c. *É sorprendente que José emendar a factura.
d. *É sorprendente José emende a factura.
"It is surprising that José amends the invoice."

(25) a. É necesario que Ana rememore o libro.
b. É necesario Ana rememorar o libro.
c. *É necesario que Ana rememorar o libro.
d. *É necesario Ana rememore o libro.
"It is necessary that Ana reminisces about the book."

(26) a. É importante que Marta valide o comunicado.
b. É importante Marta validar o comunicado.
c. *É importante que Marta validar o comunicado.
d. *É importante Marta valide o comunicado.
"It is important that Marta validates the statement."

(27) a. É posible que Laura falsee a proposta.
b. É posible Laura falsear a proposta.
c. *É posible que Laura falsear a proposta.
d. *É posible Laura falsee a proposta.
"It is possible that Laura falsifies the proposal."

(28) a. É probable que Francisco tramite a lei.
b. É probable Francisco tramitar a lei.
c. *É probable que Francisco tramitar a lei.
d. *É probable Francisco tramite a lei.
"It is probable that Francisco processes the law."

(29) a. É sorprendente que Victoria abrevie a carta.
b. É sorprendente Victoria abreviar a carta.
c. *É sorprendente que Victoria abreviar a carta.
d. *É sorprendente Victoria abrevie a carta.
"It is surprising that Victoria abbreviates the letter."

(30) a. É necesario que Juan rectifique o informe.
b. É necesario Juan rectificar o informe.
c. *É necesario que Juan rectificar o informe.
d. *É necesario Juan rectifique o informe.
"It is necessary that Juan rectifies the report."

(31) a. É importante que Salvador exalte o xardín.
b. É importante Salvador exaltar o xardín.
c. *É importante que Salvador exaltar o xardín.
d. *É importante Salvador exalte o xardín.
"It is important that Salvador exalts the garden."

(32) a. É posible que Alejandro amoble o almacén.
b. É posible Alejandro amoblar o almacén.
c. *É posible que Alejandro amoblar o almacén.
d. *É posible Alejandro amoble o almacén.
"It is possible that Alejandro furnishes the warehouse."

(33) a. É probable que Elena ventile o apartamento.
b. É probable Elena ventilar o apartamento.
c. *É probable que Elena ventilar o apartamento.
d. *É probable Elena ventile o apartamento.
"It is probable that Elena ventilates the apartment."

(34) a. É sorprendente que Javier desaloxe a casa.
b. É sorprendente Javier desaloxar a casa.
c. *É sorprendente que Javier desaloxar a casa.
d. *É sorprendente Javier desaloxe a casa.
"It is surprising that Javier vacates the house."

(35) a. É necesario que Sara fregue a habitación.
b. É necesario Sara fregar a habitación.
c. *É necesario que Sara fregar a habitación.
d. *É necesario Sara fregue a habitación.
"It is necessary that Sara mops the room."

(36) a. É importante que Alba embeleza a escultura.
b. É importante Alba embelecer a escultura.
c. *É importante que Alba embelecer a escultura.
d. *É importante Alba embeleza a escultura.
"It is important that Alba beautifies the sculpture."

(37) a. É posible que Ángela arrende a cociña.
b. É posible Ángela arrendar a cociña.
c. *É posible que Ángela arrendar a cociña.
d. *É posible Ángela arrende a cociña.
"It is possible that Ángela leases the kitchen."

(38) a. É probable que Carlos eloxie o lago.
b. É probable Carlos eloxiar o lago.
c. *É probable que Carlos eloxiar o lago.
d. *É probable Carlos eloxie o lago.
"It is probable that Carlos praises the lake."

(39) a. É sorprendente que Julia desatenda a paisaxe.
b. É sorprendente Julia desatender a paisaxe.
c. *É sorprendente que Julia desatender a paisaxe.
d. *É sorprendente Julia desatenda a paisaxe.
"It is surprising that Julia neglects the landscape."

(40) a. É necesario que Miguel custodie o cadro.
b. É necesario Miguel custodiar o cadro.
c. *É necesario que Miguel custodiar o cadro.
d. *É necesario Miguel custodie o cadro.
"It is necessary that Miguel guards the painting."

(41) a. É importante que Ana emende a lei.
b. É importante Ana emendar a lei.
c. *É importante que Ana emendar a lei.
d. *É importante Ana emende a lei.
"It is important that Ana amends the law."

(42) a. É posible que Manuel rememore a carta.
b. É posible Manuel rememorar a carta.
c. *É posible que Manuel rememorar a carta.
d. *É posible Manuel rememore a carta.
"It is possible that Manuel reminisces about the letter."

(43) a. É probable que Antonio simplifique o libro.
b. É probable Antonio simplificar o libro.
c. *É probable que Antonio simplificar o libro.
d. *É probable Antonio simplifique o libro.
"It is probable that Antonio simplifies the book."

(44) a. É sorprendente que Sofía escanee a postal.
b. É sorprendente Sofía escanear a postal.
c. *É sorprendente que Sofía escanear a postal.
d. *É sorprendente Sofía escanee a postal.
"It is surprising that Sofía scans the postcard."

(45) a. É necesario que José transcriba o correo.
b. É necesario José transcribir o correo.
c. *É necesario que José transcribir o correo.
d. *É necesario José transcriba o correo.
"It is necessary that José transcribes the email."

(46) a. É importante que Juan abrevie a novela.
b. É importante Juan abreviar a novela.
c. *É importante que Juan abreviar a novela.
d. *É importante Juan abrevie a novela.
"It is important that Juan abbreviates the novel."

(47) a. É posible que Marta rectifique a factura.
b. É posible Marta rectificar a factura.
c. *É posible que Marta rectificar a factura.
d. *É posible Marta rectifique a factura.
"It is possible that Marta rectifies the invoice."

(48) a. É probable que Laura valide o informe.
b. É probable Laura validar o informe.
c. *É probable que Laura validar o informe.
d. *É probable Laura valide o informe.
"It is probable that Laura validates the report."

(49) a. É sorprendente que Francisco falsee o comunicado.
b. É sorprendente Francisco falsear o comunicado.
c. *É sorprendente que Francisco falsear o comunicado.
d. *É sorprendente Francisco falsee o comunicado.
"It is surprising that Francisco falsifies the statement."

(50) a. É necesario que Victoria tramite a proposta.
b. É necesario Victoria tramitar a proposta.
c. *É necesario que Victoria tramitar a proposta.
d. *É necesario Victoria tramite a proposta.
"It is necessary that Victoria processes the proposal."

(51) a. É importante que Sara desaloxe o apartamento.
b. É importante Sara desaloxar o apartamento.
c. *É importante que Sara desaloxar o apartamento.
d. *É importante Sara desaloxe o apartamento.
"It is important that Sara vacates the apartment."

(52) a. É posible que Salvador fregue a casa.
b. É posible Salvador fregar a casa.
c. *É posible que Salvador fregar a casa.
d. *É posible Salvador fregue a casa.
"It is possible that Salvador mops the house."

(53) a. É probable que Alejandro exalte a escultura.
b. É probable Alejandro exaltar a escultura.
c. *É probable que Alejandro exaltar a escultura.
d. *É probable Alejandro exalte a escultura.
"It is probable that Alejandro exalts the sculpture."

(54) a. É sorprendente que Elena amoble a cociña.
b. É sorprendente Elena amoblar a cociña.
c. *É sorprendente que Elena amoblar a cociña.
d. *É sorprendente Elena amoble a cociña.
"It is surprising that Elena furnishes the kitchen."

(55) a. É necesario que Javier ventile o almacén.
b. É necesario Javier ventilar o almacén.
c. *É necesario que Javier ventilar o almacén.
d. *É necesario Javier ventile o almacén.
"It is necessary that Javier ventilates the warehouse."

(56) a. É importante que Miguel desatenda o lago.
b. É importante Miguel desatender o lago.
c. *É importante que Miguel desatender o lago.
d. *É importante Miguel desatenda o lago.
"It is important that Miguel neglects the lake."

(57) a. É posible que Alba custodie a paisaxe.
b. É posible Alba custodiar a paisaxe.
c. *É posible que Alba custodiar a paisaxe.
d. *É posible Alba custodie a paisaxe.
"It is possible that Alba guards the landscape."

(58) a. É probable que Ángela embeleza o cadro.
b. É probable Ángela embelecer o cadro.
c. *É probable que Ángela embelecer o cadro.
d. *É probable Ángela embeleza o cadro.
"It is probable that Ángela beautifies the painting."

(59) a. É sorprendente que Carlos arrende a habitación.
b. É sorprendente Carlos arrendar a habitación.
c. *É sorprendente que Carlos arrendar a habitación.
d. *É sorprendente Carlos arrende a habitación.
"It is surprising that Carlos leases the room."

(60) a. É necesario que Julia eloxie o xardín.
b. É necesario Julia eloxiar o xardín.
c. *É necesario que Julia eloxiar o xardín.
d. *É necesario Julia eloxie o xardín.
"It is necessary that Julia praises the garden."

(61) a. É importante que José escanee o libro.
b. É importante José escanear o libro.
c. *É importante que José escanear o libro.
d. *É importante José escanee o libro.
"It is important that José scans the book."

(62) a. É posible que Ana transcriba a postal.
b. É posible Ana transcribir a postal.
c. *É posible que Ana transcribir a postal.
d. *É posible Ana transcriba a postal.
"It is possible that Ana transcribes the postcard."

(63) a. É probable que Manuel emende a proposta.
b. É probable Manuel emendar a proposta.
c. *É probable que Manuel emendar a proposta.
d. *É probable Manuel emende a proposta.
"It is probable that Manuel amends the proposal."

(64) a. É sorprendente que Antonio rememore a novela.
b. É sorprendente Antonio rememorar a novela.
c. *É sorprendente que Antonio rememorar a novela.
d. *É sorprendente Antonio rememore a novela.
"It is surprising that Antonio reminisces about the novel."

(65) a. É necesario que Sofía simplifique a carta.
b. É necesario Sofía simplificar a carta.
c. *É necesario que Sofía simplificar a carta.
d. *É necesario Sofía simplifique a carta.
"It is necessary that Sofía simplifies the letter."

(66) a. É importante que Victoria falsee o informe.
b. É importante Victoria falsear o informe.
c. *É importante que Victoria falsear o informe.
d. *É importante Victoria falsee o informe.
"It is important that Victoria falsifies the report."

(67) a. É posible que Juan tramite o comunicado.
b. É posible Juan tramitar o comunicado.
c. *É posible que Juan tramitar o comunicado.
d. *É posible Juan tramite o comunicado.
"It is possible that Juan processes the statement."

(68) a. É probable que Marta abrevie o correo.
b. É probable Marta abreviar o correo.
c. *É probable que Marta abreviar o correo.
d. *É probable Marta abrevie o correo.
"It is probable that Marta abbreviates the email."

(69) a. É importante que Laura rectifique a lei.
b. É importante Laura rectificar a lei.
c. *É importante que Laura rectificar a lei.
d. *É importante Laura rectifique a lei.
"It is important that Laura rectifies the law."

(70) a. É necesario que Francisco valide a factura.
b. É necesario Francisco validar a factura.
c. *É necesario que Francisco validar a factura.
d. *É necesario Francisco valide a factura.
"It is necessary that Francisco validates the invoice."

(71) a. É importante que Javier amoble a habitación.
b. É importante Javier amoblar a habitación.
c. *É importante que Javier amoblar a habitación.
d. *É importante Javier amoble a habitación.
"It is important that Javier furnishes the room."

(72) a. É posible que Sara ventile a cociña.
b. É posible Sara ventilar a cociña.
c. *É posible que Sara ventilar a cociña.
d. *É posible Sara ventile a cociña.
"It is possible that Sara ventilates the kitchen."

(73) a. É probable que Salvador desaloxe o almacén.
b. É probable Salvador desaloxar o almacén.
c. *É probable que Salvador desaloxar o almacén.
d. *É probable Salvador desaloxe o almacén.
"It is probable that Salvador vacates the warehouse."

(74) a. É sorprendente que Alejandro fregue o apartamento.
b. É sorprendente Alejandro fregar o apartamento.
c. *É sorprendente que Alejandro fregar o apartamento.
d. *É sorprendente Alejandro fregue o apartamento.
"It is surprising that Alejandro mops the apartment."

(75) a. É necesario que Elena exalte o cadro.
b. É necesario Elena exaltar o cadro.
c. *É necesario que Elena exaltar o cadro.
d. *É necesario Elena exalte o cadro.
"It is necessary that Elena exalts the painting."

(76) a. É importante que Julia arrende a casa.
b. É importante Julia arrendar a casa.
c. *É importante que Julia arrendar a casa.
d. *É importante Julia arrende a casa.
"It is important that Julia leases the house."

(77) a. É posible que Miguel eloxie a escultura.
b. É posible Miguel eloxiar a escultura.
c. *É posible que Miguel eloxiar a escultura.
d. *É posible Miguel eloxie a escultura.
"It is possible that Miguel praises the sculpture."

(78) a. É probable que Alba desatenda o xardín.
b. É probable Alba desatender o xardín.
c. *É probable que Alba desatender o xardín.
d. *É probable Alba desatenda o xardín.
"It is probable that Alba neglects the garden."

(79) a. É sorprendente que Ángela custodie o lago.
b. É sorprendente Ángela custodiar o lago.
c. *É sorprendente que Ángela custodiar o lago.
d. *É sorprendente Ángela custodie o lago.
"It is surprising that Ángela guards the lake."

(80) a. É necesario que Carlos embeleza a paisaxe.
b. É necesario Carlos embelecer a paisaxe.
c. *É necesario que Carlos embelecer a paisaxe.
d. *É necesario Carlos embeleza a paisaxe.
"It is necessary that Carlos beautifies the landscape."

6. Vocabulary test used in Experiment 3

This is the "Test of Spanish verbs" participants completed at the end of Experiment 3. Participants were given a list of verbs and had to indicate if they knew the meaning of each verb by ticking the cell next to each word. At the end of the test, participants were asked some questions regarding the verbs they reported not knowing the meaning of. For each verb, they had to indicate if they had heard it before (yes/no) and if they would be able to interpret it if it appeared in a sentence (yes/no).

TEST DE VERBOS DEL ESPAÑOL

Hola, este es un test de vocabulario en español. En esta página encontrarás una serie de verbos. Por favor, indica si conoces el significado de cada verbo marcando la celda al lado de cada palabra. Por ejemplo, si conoces el verbo "comer" lo puedes indicar de esta manera:

Verbo	**¿Lo conoces?**
Comer	√

¡Muchísimas gracias por adelantado!

Verbo	**¿Lo conoces?**
Desmontar	
Rifar	
Silenciar	
Reciclar	
Enchufar	
Degustar	
Engullir	
Humedecer	
Recubrir	
Destapar	
Posponer	
Arbitrar	
Costear	
Invalidar	
Atrasar	
Embalar	
Obsequiar	
Ofertar	
Abaratar	
Acortar	
Transcribir	
Abreviar	

Rememorar	
Simplificar	
Escanear	
Tramitar	
Enmendar	
Rectificar	
Validar	
Falsear	
Ventilar	
Desalojar	
Fregar	
Arrendar	
Amueblar	
Elogiar	
Desatender	
Custodiar	
Embellecer	
Exaltar	

Si has respondido que no conocías alguno de los verbos anteriores, indícanos para cada verbo si lo habías escuchado antes y si podrías interpretarlo si apareciera en una oración.

Nombre del verbo:

¿Habías escuchado este verbo antes? SÍ NO

¿Crees que podrías interpretarlo si apareciera en una oración? SÍ NO

Nombre del verbo:

¿Habías escuchado este verbo antes? SÍ NO

¿Crees que podrías interpretarlo si apareciera en una oración? SÍ NO

Si has respondido que no conocías algún otro verbo, por favor, indica a continuación cuál es ese verbo, si lo habías escuchado antes y si podrías interpretarlo si apareciera en una oración:

7. Rubric to evaluate awareness and transcription of verbal reports in Experiment 3

Rubric evaluating awareness

The questions asked in the verbal report were:

1. As mentioned during the experiment, the sentences you listened to were formed according to two structures. Did you notice which structures were these?
2. If yes, please indicate which you think these structures are.

I present below the rubric used to evaluate awareness of the structures based on answers to these questions.

Participant status	Description
Aware	The participant states that a structure contains the complementizer *que* and then a verb conjugated (in the present subjunctive). Additionally or alternatively, the participant reports that the other structure does not contain the complementizer *que* and contains a verb in the infinitive.
Unaware	The participant is not able to identify the varying elements in the structures, i.e. the presence/absence of the complementizer and the finite or non-finite nature of the embedded verb. Alternatively, s/he is able to identify them but cannot appropriately correlate them.

TABLE B-7.1. Rubric used to classify participants as *aware* or *unaware* of the L2 structures based on their responses in the verbal report of Experiment 3 (same as in Experiment 2).

Verbal reports and awareness

Participant	Report	Awareness
P1	Frases con *que* > verbo de persona conjugado: *Es importante que Laura recomiende el libro*. Frases sin *que* > verbo no conjugado: *Es importante Laura regar las plantas.*	Aware
P2	La estructura que contenía un *que* conjugaba el verbo.	Aware
P3	Estructuras de relativo con *que* o dos estructuras con diferencia en la preposición o con cambios en el género del sustantivo que sigue a la preposición.	Unaware
P4	*Que* más verbo conjugado. Si no pone *que*, entonces el verbo iría en infinitivo.	Aware
P5	Hay estructuras que incluyen *que* y otras no. También hay estructuras en infinitivo y subjuntivo como en castellano. Algunas estructuras parecen parte de una conversación.	Unaware

Participant	Report	Awareness
P6	Al introducir subordinadas el *que*, el verbo se conjuga, mientras que en la otra estructura, la misma idea puede expresarse sin el *que*, en cuyo caso el verbo ha de ir en infinitivo.	Aware
P7	Hay dos estructuras. En una de ellas, cuando aparece el *que*, el verbo no tiene que estar en infinitivo. En la segunda estructura, cuando no aparece el *que*, el verbo tiene que estar en infinitivo.	Aware
P8	Había dos tipos de oraciones: Las que llevaban un *que* y las que no. Ambas podían ser correctas dependiendo del verbo y del complemento directo que lo acompañaba. No me ha quedado claro por qué, pero había oraciones que eran correctas cuando llevaban el *que* y cuando no lo llevaban.	Unaware
P9	No he encontrado las dos estructuras.	Unaware
P10	Siempre que antes del sujeto de la frase había un *que*, el siguiente verbo estaba conjugado. Cuando el sujeto no iba detrás de un *que*, el verbo iba sin conjugar.	Aware
P11	Creo que en la primera parte, cuando empieza la oración, sería algo para tener en cuenta si va luego un *que* o no va.	Unaware
P12	Están estructuradas por un *Es probable/posible…*, un sujeto/nombre de persona y luego un verbo seguido de *a* u *o*.	Unaware
P13	Subordinadas con *que* y el verbo conjugado y otras sin *que* y el verbo en infinitivo.	Aware
P14	Hay estructuras que presentan *que* y otras que no.	Unaware
P15	Cuando aparece *que* el verbo está conjugado y cuando no aparece el verbo aparece en infinitivo.	Aware
P16	Al principio he pensado que dependiendo del adjetivo, la frase iba con *que* o sin *que*, pero luego he visto que no.	Unaware
P17	*Que* va con verbo conjugado y, si no es oración con *que*, el verbo está en infinitivo.	Aware
P18	Las dos estructuras son *que* y lo que en castellano conocemos como el verbo en infinitivo. Había oraciones con objeto directo e indirecto. Las del *que*, por ejemplo: *Es importante que saques la basura* y otras en las que solo había objeto directo y en las que se encontraba el verbo en infinitivo, por ejemplo: *Es importante sacar la basura*.	Aware
P19	Me he fijado que algunas frases tenían un *que* y otras no, pero no sé muy bien por qué.	Unaware
P20	Una de las estructuras tiene un *que* y la otra no, pero independientemente de que esté ese *que* o no, se coloca *a* u *o*.	Unaware
P21	-	Unaware

Participant	**Report**	**Awareness**
P22	La estructura con el *que* delante del sujeto lleva la forma del verbo en presente o cualquier tiempo verbal, sin embargo, en la estructura que comienza por el sujeto, el verbo de la oración está en infinitivo.	Aware
P23	La primera [estructura] cuenta con un *que* y el verbo que sigue después del nombre es un subjuntivo. La segunda [estructura] no lleva *que* y el verbo que sigue después del nombre es un infinitivo.	Aware
P24	Cuando la frase tiene un *que*, el verbo cambia, se conjuga. Cuando la frase no tiene un *que*, el verbo se queda normal, sin conjugarse, acabando en *–ar, –er, –ir.*	Aware
P25	Cuando se utiliza *que,* el verbo se ajusta al sujeto. En cambio cuando no aparece el *que*, los verbos no se refieren a nadie, simplemente son verbos.	Aware
P26	He pensado que las estructuras podían tener algo que ver con que hubiera un artículo determinado o indeterminado en la frase.	Unaware
P27	En algunas frases había un *que* y en otras no.	Unaware
P28	La diferencia que he encontrado es que las oraciones que contienen un *que* tienen después el verbo conjugado, mientras que las oraciones que no lo contienen tienen el verbo en infinitivo.	Aware
P29	Al principio me fijaba en si las estructuras tenían que ver con la *o* u la *a* que aparecía delante del último nombre, pero al final he visto que no.	Unaware
P30	Es + adjetivo + *que* + persona nombre + verbo conjugado. Es + adjetivo + persona nombre + verbo en infinitivo.	Aware
P31	A veces las frases eran más cortas, otras veces se alargaban, pero no sabría decir por qué.	Unaware
P32	Pensaba que quizá las estructuras tenían algo que ver con las palabras *a* y *o* delante de los nombres, pero al final me ha parecido que no era correcto.	Unaware
P33	Por un lado, estaban las oraciones que tenían la palabra *que*. Estas tenían luego el verbo lo que parecía ser conjugado. Por otro lado, en las oraciones sin la palabra *que*, el verbo estaba en infinitivo.	Aware
P34	Las oraciones con el verbo terminado en *–ar, –er, –ir, –or, –ur* no tienen la palabra *que* por delante. El resto de verbos tienen un *que* por delante.	Aware
P35	He pensado que las dos estructuras podrían referirse a si la persona a la que se habla es masculino o femenino, a si los nombres iban precedidos de *o* u *a*, o a si en función del significado del verbo la frase tenía *que* o no.	Unaware

Participant	Report	Awareness
P36	He intentado adivinar si, dependiendo del significado del verbo, iba el *que* o no.	Unaware
P37	No sé si se refiere a que algunas oraciones están unidas con *a* y otras con *o*. Si no es así, no me he dado cuenta de cuáles eran las dos estructuras.	Unaware
P38	He visto que a veces la oración tenía *que* y a veces no. También he visto que a veces los verbos iban en infinitivo y otras no. He intentado ver si esto dependía de que el artículo de la frase fuese *a* u *o*.	Unaware
P39	Me he dado cuenta de que a veces las oraciones tenían un *que* y otras no y de que a veces el verbo terminaba en *–ar, –er, –ir* y otras no.	Unaware
P40	Estructura 1: Verbo principal + sujeto + verbo subordinado en infinitivo + complemento directo. Estructura 2: Verbo principal + partícula *que* + sujeto + verbo subordinado conjugado + complemento directo.	Aware
P41	La diferencia es que el primer modelo de oración (el sin *que*) tiene el verbo en infinitivo, mientras que en el segundo modelo, el verbo se conjuga.	Aware
P42	Una de las estructuras incluía *que* para introducir el subjuntivo y la otra utilizaba el infinitivo para sustituir el *que* más subjuntivo presente.	Aware
P43	Primero he pensado que quizá las estructuras tenían que ver con si la frase tenía complemento directo o indirecto. Luego he pensado que la palabra que venía antes del *que* podía tener algo que ver. También he mirado si las estructuras correctas dependían de que el artículo de la frase fuese *o* u *a*.	Unaware
P44	Una estructura tenía *que* y otra no.	Unaware

TABLE B-7.2. Transcription of participants' responses in the verbal report of Experiment 3 (in Spanish). Based on these responses, participants are classified as *aware* or *unaware* of the L2 structures.

English translation of verbal reports

P1: Sentences with *that* > conjugated verb: *It is important that Laura recommends the book*. Sentences without *that* > unconjugated verb: *It is important Laura to water the plants.*

P2: The structure that contained *that* conjugated the verb.

P3: Relative structures with *that*, or two structures with a difference in the preposition or with changes in the gender of the noun that follows the preposition.

P4: *That* plus conjugated verb. If it [the structure] does not have *that*, then the verb is an infinitive.

P5: There are structures that include *that* and others that do not. There are also structures in the infinitive and subjunctive as in Spanish. Some structures seem part of a conversation.

P6: When introducing subordinate clauses with *that,* the verb is conjugated, while, in the other structure, the same idea can be expressed without *that*, in which case the verb must be in the infinitive.

P7: There are two structures. In one of them appears *that* and the verb does not have to be in the infinitive. In the second structure, *that* does not appear and the verb has to be in the infinitive.

P8: There were two types of sentences: Those that had *that* and those that did not. Both could be correct depending on the verb and the direct object that accompanied them. It is not clear to me why, but some sentences were correct when they included *that* and when they did not.

P9: I have not found the structures.

P10: Whenever there was *that* before the subject of the sentence, the following verb was conjugated. When the subject did not follow *that*, the verb was unconjugated.

P11: I think that something to take into account is whether *that* goes after the beginning of the sentence or not.

P12: They [sentences] are structured as *It is probable/possible…*, a subject/person's name and then a verb followed by *a* or *o*.

P13: Subordinate clauses with *that* and a conjugated verb; clauses without *that* and an infinitive verb.

P14: There are structures that include *that* and others that do not.

P15: When *that* appears, the verb is conjugated and when it does not appear, the verb is in the infinitive.

P16: At first, I thought that depending on the adjective, the sentence had *that* or not, but then I saw that this was not true.

P17: *That* is followed by a conjugated verb. If the sentence does not have *that*, the verb is in the infinitive.

P18: The two structures are *that* and what we know in Spanish as the infinitive verb. There were sentences with *that*, for example: *It is important that you take out the trash* and others in which there was only an infinitive verb, for example: *It is important to take out the trash*.

P19: I noticed that some sentences had *that* and others did not, but I do not really know why.

P20: One of the structures has *that* and the other does not, but regardless of whether *that* is present or not, the sentences have *a* or *o*.

P22: The structure that has *that* in front of the subject contains a verb in the present or in any tense. However, in the structure that begins with the subject, the verb of the sentence is in the infinitive.

P23: The first [structure] has *that* and the verb that follows the noun is a subjunctive. The second [structure] does not have *that* and the verb that follows the noun is an infinitive.

P24: When the sentence has *that*, the verb changes, it is conjugated. When the sentence does not have *that*, the verb remains normal, without conjugation, ending in *–ar, –er, –ir.*

P25: When *that* is used, the verb agrees with the subject. However, when *that* is not used, the verb does not refer to anyone.

P26: I thought that the structures had something to do with whether there was a definite or indefinite article in the sentence.

P27: In some sentences, there was *that* and in others there was not.

P28: The difference I have found is that sentences that contain *that* have a conjugated verb, while sentences that do not contain it have a verb in the infinitive.

P29: At first, I paid attention to whether the structures had to do with the *o* or the *a* that appeared before the last noun, but in the end I realized that they did not.

P30: It is + adjective + *that* + person's name + conjugated verb / It is + adjective + person's name + verb in infinitive.

P31: Sometimes sentences were shorter and sometimes they were longer, but I cannot say why.

P32: I thought that perhaps the structures had something to do with the words *a* and *o* in front of the nouns, but in the end I thought this was not correct.

P33: On the one hand, there were sentences that had the word *that*. These then had the verb conjugated. On the other hand, in sentences without the word *that*, the verb was in the infinitive.

P34: Sentences with the verb ending in *–ar, –er, –ir, –or, –ur* do not have the word *that*. The rest of the verbs have *that* before them.

P35: I thought that the two structures could refer to: the person spoken to being male or female, the nouns being preceded by *o* or *a,* or to the fact that depending on the meaning of the verb the sentence had *that* or not.

P36: I have tried to guess whether, depending on the meaning of the verb, *that* was included or not.

P37: I do not know if the two structures refer to the fact that some sentences have *a* and others *o*. If not, I have not realized which were the two structures.

P38: I have seen that sometimes the sentence had *that* and sometimes it did not. I have also seen that sometimes the verbs were in the infinitive and sometimes they were not. I tried to see if this depended on whether the article in the sentence was *a* or *o*.

P39: I have noticed that sometimes sentences had *that* and sometimes they did not, and that sometimes the verb ended in *–ar, –er, –ir* and sometimes it did not.

P40: Structure 1: Main verb + subject + subordinate verb in the infinitive + direct object. Structure 2: Main verb + particle *that* + subject + conjugated subordinate verb + direct object.

P41: The difference is that the first sentence model (the one without *that*) has the verb in the infinitive, while in the second model, the verb is conjugated.

P42: One of the structures used *that* to introduce the subjunctive and the other used the infinitive to replace the present subjunctive and *that*.

P43: First, I thought that perhaps the structures had to do with whether the sentence had a direct or an indirect object. Then, I thought that the word that came before *that* could have something to do with them. I have also thought that whether the structures were correct or not depended on whether the article in the sentence was *o* or *a*.

P44: One structure had *that* and the other did not.

Appendix C

Appendices to Chapter 4

1. Linguistic questionnaire used in Experiments 4 and 5

English translation of the questionnaire (original in Spanish)

Personal information		
Name		
Surname		
Sex	Male	Female
Date of birth (year)		
Email		
Phone number		
Place of residence (city and region)		
Place of birth (city and region)		
Have you ever lived in another city and/or region?	Where?	When?

Native language
When you were little, which language did you use with…
…your mother?
…your father?
…your siblings?
…your grandparents?

Age of acquisition
How old were you when you started…
…speaking in Spanish?
…speaking in Catalan?
If you have a certificate in Catalan, name which:

Language use			
Which language and how often did you use it…			
…when you were little, before starting school?			
	At school/university/work	At home	Other places
…in your childhood, at primary school?			
… in your puberty, at high school?			
… nowadays, as an adult?			
Which language do you feel most comfortable using?	Spanish	Catalan	Both

Proficiency (self-assessment)					
Rank your skills in the following languages:					
	Speaking	Listening	Reading	Writing	Certificate (if any)
Catalan					
Spanish					
English					
French					
Other					

Response options and scoring

Native language	
Spanish only	1
Mostly Spanish, rarely Catalan	2
Mostly Spanish, but Catalan at least 25% of the time	3
Spanish and Catalan with equal frequency	4
Mostly Catalan, but Spanish at least 25% of the time	5
Mostly Catalan, rarely Spanish	6
Catalan only	7

Language use	
Spanish only	1
Mostly Spanish, rarely Catalan	2
Mostly Spanish, but Catalan at least 25% of the time	3
Spanish and Catalan with equal frequency	4
Mostly Catalan, but Spanish at least 25% of the time	5
Mostly Catalan, rarely Spanish	6
Catalan only	7

Proficiency (self-assessment)	
Very poor	1
Poor	2
Enough	3
Pretty good	4
Good	5
Very good	6
Perfect	7

2. Linguistic information about the participants in Experiments 4 and 5

The tables below report the information obtained in the linguistic background questionnaire that participants in Experiments 4 and 5 filled out before the experiment. For each experiment, I report first the information about cognate and non-cognate learners' language use in different life periods and in different environments. Then, I report cognate and non-cognate learners' self-assessed proficiency in Spanish.

Experiment 4 (n = 60, 30 cognate learners and 30 non-cognate learners)

		Cognate learners	**Non-cognate learners**	**Two-sample t-tests**
Childhood	Primary school	4.43 (1.79)	3.93 (1.84)	$t(58) = 1.07, p = .29, d = 0.28$
	Home	1.73 (0.45)	2.00 (0.91)	$t(58) = -1.44, p = .16, d = -0.37$
	Other places	2.40 (0.81)	2.53 (1.36)	$t(58) = -0.46, p = .65, d = -0.12$
	Mean	2.86 (1.63)	2.82 (1.63)	$t(178) = 0.14, p = .89, d = 0.02$
Puberty	High school	4.07 (1.78)	3.60 (1.81)	$t(58) = 1.01, p = .32, d = 0.26$
	Home	1.87 (0.63)	1.90 (0.92)	$t(58) = -0.16, p = .87, d = -0.04$
	Other places	2.47 (0.73)	2.43 (1.25)	$t(58) = 0.13, p = .90, d = 0.03$
	Mean	2.80 (1.49)	2.64 (1.54)	$t(178) = 0.69, p = .49, d = 0.10$
Adulthood	University/work	3.93 (1.72)	3.50 (1.50)	$t(58) = 1.04, p = .30, d = 0.27$
	Home	1.73 (0.45)	1.80 (0.87)	$t(58) = -0.37, p = .72, d = -0.09$
	Other places	2.37 (0.61)	2.33 (0.84)	$t(58) = 0.17, p = .86, d = 0.05$
	Mean	2.68 (1.42)	2.54 (1.32)	$t(178) = 0.65, p = .51, d = 0.10$

TABLE C-2.1. Language use (SD in brackets) during childhood, puberty and adulthood in different environments as self-assessed by cognate and non-cognate learners in Experiment 4. Scores are on a 7-point scale: 1 = Spanish only; 2 = Mostly Spanish, rarely Catalan; 3 = Mostly Spanish, but Catalan at least 25% of the time; 4 = Spanish and Catalan with equal frequency; 5 = Mostly Catalan, but Spanish at least 25% of the time; 6 = Mostly Catalan, rarely Spanish; 7 = Catalan only. Scores for cognate and non-cognate learners are compared by independent-samples t-tests. Cohen's d is reported as a standardised measure of effect size.

	Cognate learners	**Non-cognate learners**	**Two-sample t-tests**
Speaking	6.70 (0.53)	6.83 (0.38)	$t(58) = -1.11$, $p = .27$, $d = -0.29$
Listening	6.87 (0.35)	6.93 (0.25)	$t(58) = -0.85$, $p = .40$, $d = -0.22$
Reading	6.80 (0.41)	6.83 (0.38)	$t(58) = -0.33$, $p = .74$, $d = -0.08$
Writing	6.63 (0.61)	6.60 (0.62)	$t(58) = 0.21$, $p = .84$, $d = 0.05$
Mean	6.75 (0.49)	6.80 (0.44)	$t(238) = -0.83$, $p = .41$, $d = -0.11$

TABLE C-2.2. Proficiency level speaking, listening, reading and writing in Spanish (SD in brackets) as self-assessed by cognate and non-cognate learners in Experiment 4. Scores are on a 7-point scale: 1 = Very poor; 2 = Poor; 3 = Enough; 4 = Pretty good; 5 = Good; 6 = Very good; 7 = Perfect. Scores for cognate and non-cognate learners are compared by independent-samples t-tests. Cohen's d is reported as a standardised measure of effect size.

Experiment 5 (n = 60, 30 cognate learners and 30 non-cognate learners)

		Cognate learners	**Non-cognate learners**	**Two-sample t-tests**
Childhood	Primary school	3.97 (1.75)	4.27 (1.93)	$t(58) = -0.63$, $p = .53$, $d = -0.16$
	Home	2.00 (0.95)	1.90 (1.18)	$t(58) = 0.36$, $p = .72$, $d = 0.09$
	Other places	2.70 (1.58)	3.03 (1.94)	$t(58) = -0.73$, $p = 47$, $d = -0.19$
	Mean	2.89 (1.67)	3.07 (1.96)	$t(178) = -0.66$, $p = .51$, $d = -0.10$
Puberty	High school	3.50 (1.57)	3.8 (1.75)	$t(58) = -0.70$, $p = .49$, $d = -0.18$
	Home	1.93 (0.91)	1.97 (1.19)	$t(58) = -0.12$, $p = .90$, $d = -0.03$
	Other places	2.43 (1.30)	2.87 (1.68)	$t(58) = -1.12$, $p = .27$, $d = -0.29$
	Mean	2.62 (1.43)	2.88 (1.71)	$t(178) = -1.08$, $p = .28$, $d = -0.16$
Adulthood	University/work	3.33 (1.35)	3.57 (1.43)	$t(58) = -0.65$, $p = .52$, $d = -0.19$
	Home	2.00 (0.91)	1.93 (1.17)	$t(58) = 0.25$, $p = .81$, $d = 0.06$
	Other places	2.50 (0.94)	2.57 (0.68)	$t(58) = -0.32$, $p = .75$, $d = -0.08$
	Mean	2.61 (1.21)	2.69 (1.31)	$t(178) = -0.42$, $p = .68$, $d = -0.06$

TABLE C-2.3. Language use (SD in brackets) during childhood, puberty and adulthood in different environments as self-assessed by cognate and non-cognate learners in Experiment 5. Scores are on a 7-point scale: 1 = Spanish only; 2 = Mostly Spanish, rarely Catalan; 3 = Mostly Spanish, but Catalan at least 25% of the time; 4 = Spanish and Catalan with equal frequency; 5 = Mostly Catalan, but Spanish at least 25% of the time; 6 = Mostly Catalan, rarely Spanish; 7 = Catalan only. Scores for cognate and non-cognate learners are compared by independent-samples t-tests. Cohen's d is reported as a standardised measure of effect size.

	Cognate learners	**Non-cognate learners**	**Two-sample t-tests**
Speaking	6.77 (0.43)	6.83 (0.38)	*t*(58) = -0.64, *p* = .53, *d* = -0.16
Listening	6.80 (0.41)	6.93 (0.25)	*t*(58) = -1.52, *p* = .13, *d* = -0.39
Reading	6.90 (0.31)	6.90 (0.31)	*t*(58) = 0, *p* = 1, *d* = 0
Writing	6.80 (0.61)	6.77 (0.50)	*t*(58) = 0.23, *p* = .82, *d* = 0.06
Mean	6.82 (0.45)	6.86 (0.37)	*t*(238) = -0.78, *p* = .44, *d* = -0.10

TABLE C-2.4. Proficiency level speaking, listening, reading and writing in Spanish (SD in brackets) as self-assessed by cognate and non-cognate learners in Experiment 5. Scores are on a 7-point scale: 1 = Very poor; 2 = Poor; 3 = Enough; 4 = Pretty good; 5 = Good; 6 = Very good; 7 = Perfect. Scores for cognate and non-cognate learners are compared by independent-samples t-tests. Cohen’s d is reported as a standardised measure of effect size.

3. Informed consent used in Experiments 4 and 5

This is the Spanish "Informed consent in comprehension tests" participants read and signed before Experiments 4 and 5. This consent provided participants with all the necessary information about the experiment so that they could decide freely and voluntarily whether they wanted to participate. The information given included: the project the study was part of, details of the Principal Investigators of the project and the person in charge of the experiment, description, aims and procedure of the study, risks and rights of the participant and policy of conservation and processing of personal data.

CONSENTIMIENTO INFORMADO EN PRUEBAS DE COMPRENSIÓN

El presente informe tiene como objetivo primordial proporcionarle toda la información necesaria sobre el experimento en el que va a participar y sobre la conservación y tratamiento de sus datos personales, con el objetivo de que pueda decidir libre y voluntariamente sobre su participación en el mismo.

Identificación del proyecto

Título del proyecto: Cross-linguistic activation effects in bilingual language processing and learning
Financiación: Ministerio de Ciencia, Innovación y Universidades
Título del estudio: Lexical and Syntactic Co-activation in L2 Syntax Learning
Código del proyecto: PGC2018-097970-B-100

Identificación del investigador principal y forma de contacto

Nombre y apellidos: Kepa Erdozia y Mikel Santesteban
Dirección: Centro de Investigación Micaela Portilla 3.2. Dept. Lingüística y Estudios Vascos. Facultad de Letras, Universidad del País Vasco (UPV/EHU)
E-mail: kepa.erdozia@ehu.eus and mikel.santesteban@ehu.eus
Teléfono: 945013650

Identificación del investigador responsable

El investigador responsable se encargará de pasar la prueba experimental y de informarle adecuadamente.
Nombre y apellidos: Noèlia Sanahuja Cobacho
Dirección: Centro de Investigación Micaela Portilla 3.2. Dept. Lingüística y Estudios Vascos (UPV/EHU)
E-mail: noelia.sanahuja@ehu.eus

DESCRIPCIÓN Y OBJETIVOS DE LA INVESTIGACIÓN

Nuestro objetivo principal es llegar a entender cómo la activación interlingüística de las dos lenguas modula el aprendizaje y el procesamiento de la L2 en el hablante bilingüe. Dentro de este objetivo general trabajaremos con la hipótesis de que la activación interlingüística guía la mayoría de los procesos cuando utilizamos una segunda lengua. Investigaremos cual es el papel de la distancia tipológica entre L1 y L2 en los efectos de interferencia interlingüística durante el aprendizaje y procesamiento de lenguaje y el papel del conocimiento sintáctico en el aprendizaje de reglas de L2.

PROPÓSITO DEL ESTUDIO
El objetivo del presente estudio es investigar cómo el cerebro procesa el lenguaje durante el aprendizaje de una segunda lengua.

PROCEDIMIENTO
El procedimiento a seguir consiste en aprender una serie de palabras y, seguidamente, escuchar y leer oraciones con dichas palabras mientras se mira la pantalla de un ordenador. A continuación, se realizarán dos tareas: una de comprensión y la otra de producción escrita, ambas basadas en las oraciones que se han escuchado. El experimento se realizará en 1 sesión y tendrá una duración máxima de 75 minutos.
Riesgos e incomodidades
Ninguno de los procedimientos representa peligro alguno para la salud o integridad física. Todas las intervenciones se llevarán a cabo con todas las medidas preventivas requeridas en la situación de Covid-19.

DERECHOS DEL PARTICIPANTE
Cláusula de voluntariedad y derecho de revocación
La información que contienen los datos personales del participante o cualquier otro dato identificativo no se proporcionará a terceros y se protegerá la privacidad de los mismos. Los resultados de este proyecto pueden llegar a publicarse en libros o revistas especializadas o pueden usarse con finalidades didácticas. La participación en este estudio es completamente voluntaria y, como tal, puede revocar el consentimiento dado en cualquier momento, sin dar explicaciones de ningún tipo y sin que ello suponga ningún perjuicio o medida en su contra. De igual forma, a criterio del investigador, usted puede ser retirado del estudio por alguna de las siguientes razones: (a) si no cumple con los requisitos mínimos que se establezcan para participar en el estudio; (b) si por cualquier motivo se interrumpe el estudio.

Cláusula sobre el derecho a tener más información sobre el proyecto
Si colabora en este estudio, una vez haya finalizado, tendrá usted a su disposición toda la información relativa a los resultados obtenidos en el mismo. Para acceder a ella, es necesario que se ponga en contacto con el investigador responsable del proyecto a través de la dirección de e-mail que consta en este documento.

PROTECCIÓN DE DATOS:
Se le informa de que de conformidad al Reglamento Europeo de Protección de Datos (UE2016/679):

- Los datos personales que se le solicitan son:
  a) Datos de carácter identificativo: DNI/NIF, NOMBRE Y APELLIDOS, DIRECCIÓN (POSTAL, ELECTRÓNICA), TELÉFONO, IMAGEN/VOZ
  b) Datos de características personales: FECHA DE NACIMIENTO, LUGAR DE NACIMIENTO, EDAD, SEXO, NACIONALIDAD, LENGUA MATERNA
  c) Datos académicos y profesionales: FORMACIÓN, TITULACIONES
- El código del tratamiento de datos es:TI0091
- El nombre del tratamiento de datos es: DATOS GOGO ELEBIDUNA-MENTE BILINGÜE
- La finalidad de este tratamiento es: CUESTIONARIO DE PERFIL LINGÜÍSTICO DE LOS PARTICIPANTES EN LOS EXPERIMENTOS PSICOLINGÜÍSTICOS DEL GRUPO DE INVESTIGACIÓN "GOGO ELEBIDUNA/MENTE BILINGÜE"
- El responsable del tratamiento de datos es la UPV/EHU:

Identidad: Universidad del País Vasco/Euskal Herriko Unibertsitatea CIF: Q4818001B Dirección postal: Barrio Sarriena s/n, 48940-Leioa (Bizkaia) Página web: www.ehu.eus Datos de contacto del Delegado de Protección de Datos: dpd@ehu.eus

- El periodo de conservación de sus datos será: Los datos se conservarán mientras no se solicite su supresión por la persona interesada y, en cualquier caso, siempre que estén abiertos los plazos de recurso y/o reclamación procedente o mientras sigan respondiendo a la finalidad para la que fueron obtenidos.
- La legitimación del tratamiento es: su consentimiento informado.
- Cesiones y transferencias internacionales de sus datos: No se cederán datos salvo previsión legal. No se efectuarán transferencias internacionales.
- Los derechos sobre sus datos son los de acceso, supresión, rectificación, oposición, limitación del tratamiento, portabilidad y olvido. Puede ejercerlos enviando su petición a dpd@ehu.eus.
- Tiene a su disposición información adicional en http://www.ehu.eus/babestu
- La información completa sobre este tratamiento está en: https://www.ehu.eus/es/web/idazkaritza-nagusia/ikerketa-datu-pertsonalen-tratamenduak

IDENTIFICACIÓN DE LA PERSONA QUE PRESTA EL CONSENTIMIENTO

Yo, ____, con DNI nº ___________________ declaro que he leído este documento y que doy mi consentimiento a participar voluntariamente en este estudio.

____ Fecha ___________________

Voluntario/a

El investigador abajo firmante declara que el participante ha recibido la información escrita y oral necesaria para garantizar que su participación pueda considerarse libre y voluntaria.

____ Fecha ___________________

Investigador/a

4. List of vocabulary and pictures used in Experiments 4 and 5

<u>Exposure and testing nouns (Spanish-Basque non-cognates)</u>

Antzezle
Sp.”actor”
Eng. “actor”

Gidari
Sp.”piloto”
Eng. “pilot”

Epaile
Sp.”árbitro”
Eng. “referee”

Margolari
Sp.”pintor”
Eng. “painter”

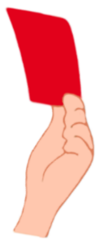
Sendagile
Sp.”médico”
Eng. “doctor”

<u>Exposure verbs (Spanish-Basque cognates/non-cognates)</u>

Pintatu / Margotu
Cognate Non-cog.
Sp.”pintar”
Eng. “paint”

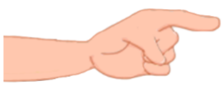
Salutatu / Agurtu
Cognate Non-cog.
Sp.”saludar”
Eng. “greet”

Presentatu/Aurkeztu
Cognate Non-cog.
Sp.”presentar”
Eng. “present”



Kastigatu/Zigortu
Cognate Non-cog.
Sp.”castigar”
Eng. “punish”

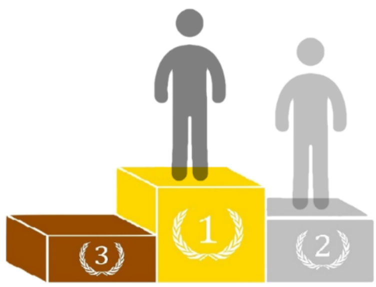
<u>Testing verbs (Spanish-Basque non-cognates)</u>



Aukeratu
Sp.”elegir”
Eng. “choose”

Aztertu
Sp.”examinar”
Eng. “examine”

Zelatatu
Sp.”espiar”
Eng. “spy on”



Gainditu
Sp.”superar”
Eng. “surpass”

5. Experimental materials used in Experiment 4

Exposure set: Sentence-picture pairs

Sentence-picture pairs constituting the exposure set in Experiment 4. Sentence-picture pairs are presented in groups of four, corresponding to the two conditions in the exposure phase in the two versions of the mini-language: a. SOV sentence-picture pair (cognate version), b. OSV sentence-picture pair (cognate version), c. SOV sentence-picture pair (non-cognate version) and d. OSV sentence-picture pair (non-cognate version). I provide the English translation shared for each group of sentences.

(1) a. Antzezleak margolaria kastigatu.
b. Margolaria antzezleak kastigatu.
c. Antzezleak margolaria zigortu.
d. Margolaria antzezleak zigortu.
"The actor is punishing the painter."

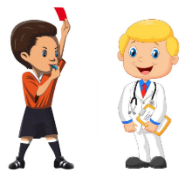
(2) a. Antzezleak sendagilea kastigatu.
b. Sendagilea antzezleak kastigatu.
c. Antzezleak sendagilea zigortu.
d. Sendagilea antzezleak zigortu.
"The actor is punishing the doctor."

(3) a. Antzezleak gidaria kastigatu.
b. Gidaria antzezleak kastigatu.
c. Antzezleak gidaria zigortu.
d. Gidaria antzezleak zigortu.
"The actor is punishing the pilot."

(4) a. Antzezleak epailea kastigatu.
b. Epailea antzezleak kastigatu.
c. Antzezleak epailea zigortu.
d. Epailea antzezleak zigortu.
"The actor is punishing the referee."

(5) a. Epaileak margolaria kastigatu.
b. Margolaria epaileak kastigatu.
c. Epaileak margolaria zigortu.
d. Margolaria epaileak zigortu.
"The referee is punishing the painter."

(6) a. Epaileak sendagilea kastigatu.
b. Sendagilea epaileak kastigatu.
c. Epaileak sendagilea zigortu.
d. Sendagilea epaileak zigortu.
"The referee is punishing the doctor."



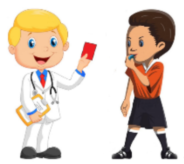
(7) a. Epaileak gidaria kastigatu.
b. Gidaria epaileak kastigatu.
c. Epaileak gidaria zigortu.
d. Gidaria epaileak zigortu.
"The referee is punishing the pilot."

(8) a. Epaileak antzezlea kastigatu.
b. Antzezlea epaileak kastigatu.
c. Epaileak antzezlea zigortu.
d. Antzezlea epaileak zigortu.
"The referee is punishing the actor."

(9) a. Sendagileak margolaria kastigatu.
b. Margolaria sendagileak kastigatu.
c. Sendagileak margolaria zigortu.
d. Margolaria sendagileak zigortu.
"The doctor is punishing the painter."

(10) a. Sendagileak gidaria kastigatu.
b. Gidaria sendagileak kastigatu.
c. Sendagileak gidaria zigortu.
d. Gidaria sendagileak zigortu.
"The doctor is punishing the pilot."

(11) a. Sendagileak epailea kastigatu.
b. Epailea sendagileak kastigatu.
c. Sendagileak epailea zigortu.
d. Epailea sendagileak zigortu.
"The doctor is punishing the referee."



(12) a. Sendagileak antzezlea kastigatu.
b. Antzezlea sendagileak kastigatu.
c. Sendagileak antzezlea zigortu.
d. Antzezlea sendagileak zigortu.
"The doctor is punishing the actor."

(13) a. Gidariak margolaria kastigatu.
b. Margolaria gidariak kastigatu.
c. Gidariak margolaria zigortu.
d. Margolaria gidariak zigortu.
"The pilot is punishing the painter."

(14) a. Gidariak sendagilea kastigatu.
b. Sendagilea gidariak kastigatu.
c. Gidariak sendagilea zigortu.
d. Sendagilea gidariak zigortu.
"The pilot is punishing the doctor."

(15) a. Gidariak epailea kastigatu.
b. Epailea gidariak kastigatu.
c. Gidariak epailea zigortu.
d. Epailea gidariak zigortu.
"The pilot is punishing the referee."

(16) a. Gidariak antzezlea kastigatu.
b. Antzezlea gidariak kastigatu.
c. Gidariak antzezlea zigortu.
d. Antzezlea gidariak zigortu.
"The pilot is punishing the actor."

(17) a. Margolariak sendagilea kastigatu.
b. Sendagilea margolariak kastigatu.
c. Margolariak sendagilea zigortu.
d. Sendagilea margolariak zigortu.
"The painter is punishing the doctor."

(18) a. Margolariak gidaria kastigatu.
b. Gidaria margolariak kastigatu.
c. Margolariak gidaria zigortu.
d. Gidaria margolariak zigortu.
"The painter is punishing the pilot."

(19) a. Margolariak epailea kastigatu.
b. Epailea margolariak kastigatu.
c. Margolariak epailea zigortu.
d. Epailea margolariak zigortu.
"The painter is punishing the referee."

(20) a. Margolariak antzezlea kastigatu.
b. Antzezlea margolariak kastigatu.
c. Margolariak antzezlea zigortu.
d. Antzezlea margolariak zigortu.
"The painter is punishing the actor."

(21) a. Antzezleak margolaria pintatu.
b. Margolaria antzezleak pintatu.
c. Antzezleak margolaria margotu.
d. Margolaria antzezleak margotu.
"The actor is painting the painter."

(22) a. Antzezleak sendagilea pintatu.
b. Sendagilea antzezleak pintatu.
c. Antzezleak sendagilea margotu.
d. Sendagilea antzezleak margotu.
"The actor is painting the doctor."

(23) a. Antzezleak gidaria pintatu.
b. Gidaria antzezleak pintatu.
c. Antzezleak gidaria margotu.
d. Gidaria antzezleak margotu.
"The actor is painting the pilot."

(24) a. Antzezleak epailea pintatu.
b. Epailea antzezleak pintatu.
c. Antzezleak epailea margotu.
d. Epailea antzezleak margotu.
"The actor is painting the referee."

(25) a. Epaileak margolaria pintatu.
b. Margolaria epaileak pintatu.
c. Epaileak margolaria margotu.
d. Margolaria epaileak margotu.
"The referee is painting the painter."

(26) a. Epaileak sendagilea pintatu.
b. Sendagilea epaileak pintatu.
c. Epaileak sendagilea margotu.
d. Sendagilea epaileak margotu.
"The referee is painting the doctor."

(27) a. Epaileak gidaria pintatu.
b. Gidaria epaileak pintatu.
c. Epaileak gidaria margotu.
d. Gidaria epaileak margotu.
"The referee is painting the pilot."

(28) a. Epaileak antzezlea pintatu.
b. Antzezlea epaileak pintatu.
c. Epaileak antzezlea margotu.
d. Antzezlea epaileak margotu.
"The referee is painting the actor."

(29) a. Sendagileak margolaria pintatu.
b. Margolaria sendagileak pintatu.
c. Sendagileak margolaria margotu.
d. Margolaria sendagileak margotu.
"The doctor is painting the painter."

(30) a. Sendagileak gidaria pintatu.
b. Gidaria sendagileak pintatu.
c. Sendagileak gidaria margotu.
d. Gidaria sendagileak margotu.
"The doctor is painting the pilot."

(31) a. Sendagileak epailea pintatu.
b. Epailea sendagileak pintatu.
c. Sendagileak epailea margotu.
d. Epailea sendagileak margotu.
"The doctor is painting the referee."

(32) a. Sendagileak antzezlea pintatu.
b. Antzezlea sendagileak pintatu.
c. Sendagileak antzezlea margotu.
d. Antzezlea sendagileak margotu.
"The doctor is painting the actor."

(33) a. Gidariak margolaria pintatu.
b. Margolaria gidariak pintatu.
c. Gidariak margolaria margotu.
d. Margolaria gidariak margotu.
"The pilot is painting the painter."

(34)
a. Gidariak sendagilea pintatu.
b. Sendagilea gidariak pintatu.
c. Gidariak sendagilea margotu.
d. Sendagilea gidariak margotu.
"The pilot is painting the doctor."

(35)
a. Gidariak epailea pintatu.
b. Epailea gidariak pintatu.
c. Gidariak epailea margotu.
d. Epailea gidariak margotu.
"The pilot is painting the referee."

(36)
a. Gidariak antzezlea pintatu.
b. Antzezlea gidariak pintatu.
c. Gidariak antzezlea margotu.
d. Antzezlea gidariak margotu.
"The pilot is painting the actor."

(37)
a. Margolariak sendagilea pintatu.
b. Sendagilea margolariak pintatu.
c. Margolariak sendagilea margotu.
d. Sendagilea margolariak margotu.
"The painter is painting the doctor."

(38)
a. Margolariak gidaria pintatu.
b. Gidaria margolariak pintatu.
c. Margolariak gidaria margotu.
d. Gidaria margolariak margotu.
"The painter is painting the pilot."

(39)
a. Margolariak epailea pintatu.
b. Epailea margolariak pintatu.
c. Margolariak epailea margotu.
d. Epailea margolariak margotu.
"The painter is painting the referee."

(40)
a. Margolariak antzezlea pintatu.
b. Antzezlea margolariak pintatu.
c. Margolariak antzezlea margotu.
d. Antzezlea margolariak margotu.
"The painter is painting the actor."

(41) a. Antzezleak margolaria presentatu.
b. Margolaria antzezleak presentatu.
c. Antzezleak margolaria aurkeztu.
d. Margolaria antzezleak aurkeztu.
"The actor is presenting the painter."

(42) a. Antzezleak sendagilea presentatu.
b. Sendagilea antzezleak presentatu.
c. Antzezleak sendagilea aurkeztu.
d. Sendagilea antzezleak aurkeztu.
"The actor is presenting the doctor."

(43) a. Antzezleak gidaria presentatu.
b. Gidaria antzezleak presentatu.
c. Antzezleak gidaria aurkeztu.
d. Gidaria antzezleak aurkeztu.
"The actor is presenting the pilot."

(44) a. Antzezleak epailea presentatu.
b. Epailea antzezleak presentatu.
c. Antzezleak epailea aurkeztu.
d. Epailea antzezleak aurkeztu.
"The actor is presenting the referee."

(45) a. Epaileak margolaria presentatu.
b. Margolaria epaileak presentatu.
c. Epaileak margolaria aurkeztu.
d. Margolaria epaileak aurkeztu.
"The referee is presenting the painter."

(46) a. Epaileak sendagilea presentatu.
b. Sendagilea epaileak presentatu.
c. Epaileak sendagilea aurkeztu.
d. Sendagilea epaileak aurkeztu.
"The referee is presenting doctor."

(47) a. Epaileak gidaria presentatu.
b. Gidaria epaileak presentatu.
c. Epaileak gidaria aurkeztu.
d. Gidaria epaileak aurkeztu.
"The referee is presenting the pilot."

(48) a. Epaileak antzezlea presentatu.
b. Antzezlea epaileak presentatu.
c. Epaileak antzezlea aurkeztu.
d. Antzezlea epaileak aurkeztu.
"The referee is presenting the actor."

(49) a. Sendagileak margolaria presentatu.
b. Margolaria sendagileak presentatu.
c. Sendagileak margolaria aurkeztu.
d. Margolaria sendagileak aurkeztu.
"The doctor is presenting the painter."

(50) a. Sendagileak gidaria presentatu.
b. Gidaria sendagileak presentatu.
c. Sendagileak gidaria aurkeztu.
d. Gidaria sendagileak aurkeztu.
"The doctor is presenting the pilot."

(51) a. Sendagileak epailea presentatu.
b. Epailea sendagileak presentatu.
c. Sendagileak epailea aurkeztu.
d. Epailea sendagileak aurkeztu.
"The doctor is presenting the referee."

(52) a. Sendagileak antzezlea presentatu.
b. Antzezlea sendagileak presentatu.
c. Sendagileak antzezlea aurkeztu.
d. Antzezlea sendagileak aurkeztu.
"The doctor is presenting the actor."

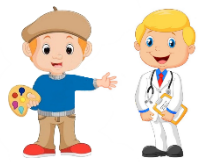
(53) a. Gidariak margolaria presentatu.
b. Margolaria gidariak presentatu.
c. Gidariak margolaria aurkeztu.
d. Margolaria gidariak aurkeztu.
"The pilot is presenting the painter."

(54) a. Gidariak sendagilea presentatu.
b. Sendagilea gidariak presentatu.
c. Gidariak sendagilea aurkeztu.
d. Sendagilea gidariak aurkeztu.
"The pilot is presenting the doctor."

(55) a. Gidariak epailea presentatu.
b. Epailea gidariak presentatu.
c. Gidariak epailea aurkeztu.
d. Epailea gidariak aurkeztu.
"The pilot is presenting the referee."

(56) a. Gidariak antzezlea presentatu.
b. Antzezlea gidariak presentatu.
c. Gidariak antzezlea aurkeztu.
d. Antzezlea gidariak aurkeztu.
"The pilot is presenting the actor."

(57) a. Margolariak sendagilea presentatu.
b. Sendagilea margolariak presentatu.
c. Margolariak sendagilea aurkeztu.
d. Sendagilea margolariak aurkeztu.
"The painter is presenting the doctor."



(58) a. Margolariak gidaria presentatu.
b. Gidaria margolariak presentatu.
c. Margolariak gidaria aurkeztu.
d. Gidaria margolariak aurkeztu.
"The painter is presenting the pilot."

(59) a. Margolariak epailea presentatu.
b. Epailea margolariak presentatu.
c. Margolariak epailea aurkeztu.
d. Epailea margolariak aurkeztu.
"The painter is presenting the referee."

(60) a. Margolariak antzezlea presentatu.
b. Antzezlea margolariak presentatu.
c. Margolariak antzezlea aurkeztu.
d. Antzezlea margolariak aurkeztu.
"The painter is presenting the actor."

(61) a. Antzezleak margolaria salutatu.
b. Margolaria antzezleak salutatu.
c. Antzezleak margolaria agurtu.
d. Margolaria antzezleak agurtu.
"The actor is greeting the painter."

(62) a. Antzezleak sendagilea salutatu.
b. Sendagilea antzezleak salutatu.
c. Antzezleak sendagilea agurtu.
d. Sendagilea antzezleak agurtu.
"The actor is greeting the doctor."

(63) a. Antzezleak gidaria salutatu.
b. Gidaria antzezleak salutatu.
c. Antzezleak gidaria agurtu.
d. Gidaria antzezleak agurtu.
"The actor is greeting the pilot."

(64) a. Antzezleak epailea salutatu.
b. Epailea antzezleak salutatu.
c. Antzezleak epailea agurtu.
d. Epailea antzezleak agurtu.
"The actor is greeting the referee."

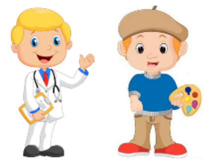
(65) a. Epaileak margolaria salutatu.
b. Margolaria epaileak salutatu.
c. Epaileak margolaria agurtu.
d. Margolaria epaileak agurtu.
"The referee is greeting the painter."

(66) a. Epaileak sendagilea salutatu.
b. Sendagilea epaileak salutatu.
c. Epaileak sendagilea agurtu.
d. Sendagilea epaileak agurtu.
"The referee is greeting the doctor."

(67) a. Epaileak gidaria salutatu.
b. Gidaria epaileak salutatu.
c. Epaileak gidaria agurtu.
d. Gidaria epaileak agurtu.
"The referee is greeting the pilot."

(68) a. Epaileak antzezlea salutatu.
b. Antzezlea epaileak salutatu.
c. Epaileak antzezlea agurtu.
d. Antzezlea epaileak agurtu.
"The referee is greeting the actor."

(69) a. Sendagileak margolaria salutatu.
b. Margolaria sendagileak salutatu.
c. Sendagileak margolaria agurtu.
d. Margolaria sendagileak agurtu.
"The doctor is greeting the painter."



(70) a. Sendagileak gidaria salutatu.
b. Gidaria sendagileak salutatu.
c. Sendagileak gidaria agurtu.
d. Gidaria sendagileak agurtu.
"The doctor is greeting the pilot."

(71) a. Sendagileak epailea salutatu.
b. Epailea sendagileak salutatu.
c. Sendagileak epailea agurtu.
d. Epailea sendagileak agurtu.
"The doctor is greeting the referee."

(72) a. Sendagileak antzezlea salutatu.
b. Antzezlea sendagileak salutatu.
c. Sendagileak antzezlea agurtu.
d. Antzezlea sendagileak agurtu.
"The doctor is greeting the actor."

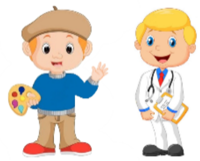
(73) a. Gidariak margolaria salutatu.
b. Margolaria gidariak salutatu.
c. Gidariak margolaria agurtu.
d. Margolaria gidariak agurtu.
"The pilot is greeting the painter."

(74) a. Gidariak sendagilea salutatu.
b. Sendagilea gidariak salutatu.
c. Gidariak sendagilea agurtu.
d. Sendagilea gidariak agurtu.
"The pilot is greeting the doctor."

(75) a. Gidariak epailea salutatu.
b. Epailea gidariak salutatu.
c. Gidariak epailea agurtu.
d. Epailea gidariak agurtu.
"The pilot is greeting the referee."

(76) a. Gidariak antzezlea salutatu.
b. Antzezlea gidariak salutatu.
c. Gidariak antzezlea agurtu.
d. Antzezlea gidariak agurtu.
"The pilot is greeting the actor."

(77) a. Margolariak sendagilea salutatu.
b. Sendagilea margolariak salutatu.
c. Margolariak sendagilea agurtu.
d. Sendagilea margolariak agurtu.
"The painter is greeting the doctor."



(78) a. Margolariak gidaria salutatu.
b. Gidaria margolariak salutatu.
c. Margolariak gidaria agurtu.
d. Gidaria margolariak agurtu.
"The painter is greeting the pilot."

(79) a. Margolariak epailea salutatu.
b. Epailea margolariak salutatu.
c. Margolariak epailea agurtu.
d. Epailea margolariak agurtu.
"The painter is greeting the referee."

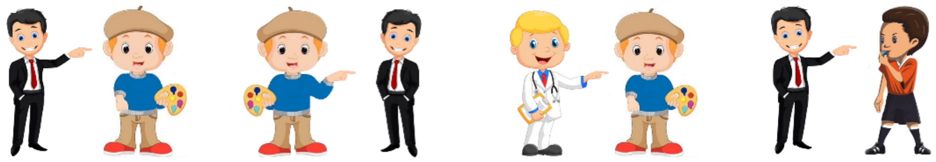
(80) a. Margolariak antzezlea salutatu.
b. Antzezlea margolariak salutatu.
c. Margolariak antzezlea agurtu.
d. Antzezlea margolariak agurtu.
"The painter is greeting the actor."

<u>Testing set: Sentence-picture pairs used in the sentence-picture congruency task</u>

Sentence-picture pairs constituting the testing set in Experiment 4. Sentences are presented in pairs: a. SOV sentence and b. OSV sentence. Each pair of sentences is associated with four pictures, resulting in the four experimental conditions in the sentence-picture congruency task: a/b-1. Syntactically congruent sentence-picture pair, a/b-2. Syntactically incongruent sentence-picture pair, a/b-3. Semantically incongruent sentence-picture pair with agent violation and a/b-4. Semantically incongruent sentence-picture pair with patient violation. I provide the English translation shared for each pair of sentences.

(1) a. Antzezleak margolaria aukeratu.
b. Margolaria antzezleak aukeratu.
"The actor is choosing the painter."

1 2 3 4

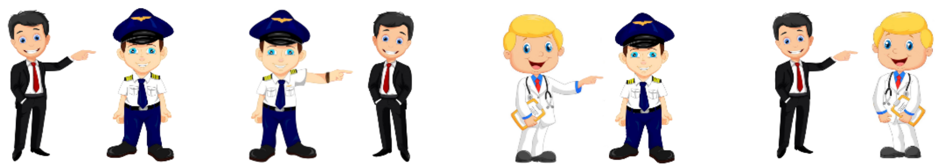


(2) a. Antzezleak sendagilea aukeratu.
b. Sendagilea antzezleak aukeratu.
"The actor is choosing the doctor."

1 2 3 4

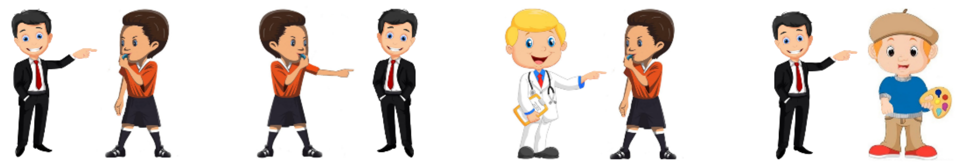
(3) a. Antzezleak gidaria aukeratu.
b. Gidaria antzezleak aukeratu.
"The actor is choosing the pilot."

1 2 3 4



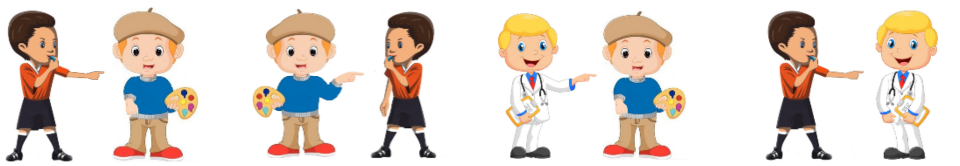
(4) a. Antzezleak epailea aukeratu.
b. Epailea antzezleak aukeratu.
"The actor is choosing the referee."

1 2 3 4



(5) a. Epaileak margolaria aukeratu.
b. Margolaria epaileak aukeratu.
"The referee is choosing the painter."

1 2 3 4

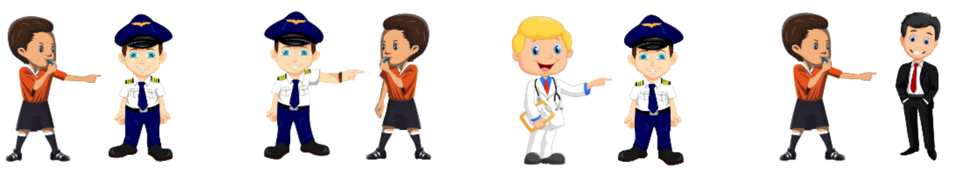


(6) a. Epaileak sendagilea aukeratu.
b. Sendagilea epaileak aukeratu.
"The referee is choosing the doctor."

1 2 3 4

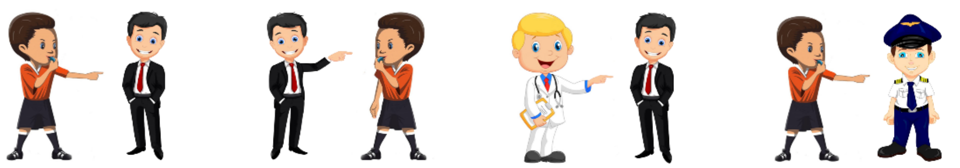
(7) a. Epaileak gidaria aukeratu.
b. Gidaria epaileak aukeratu.
"The referee is choosing the pilot."

1 2 3 4



(8) a. Epaileak antzezlea aukeratu.
b. Antzezlea epaileak aukeratu.
"The referee is choosing the actor."

1 2 3 4

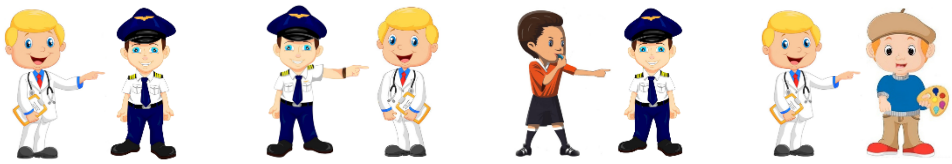


(9) a. Sendagileak margolaria aukeratu.
b. Margolaria sendagileak aukeratu.
"The doctor is choosing the painter."

1 2 3 4

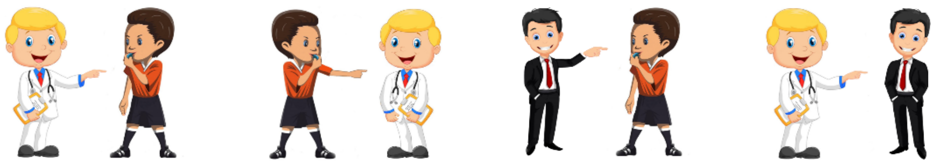
(10) a. Sendagileak gidaria aukeratu.
b. Gidaria sendagileak aukeratu.
"The doctor is choosing the pilot."

1 2 3 4



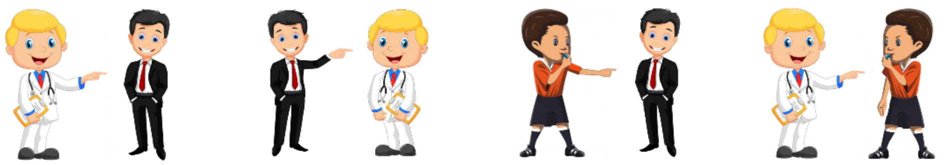
(11) a. Sendagileak epailea aukeratu.
b. Epailea sendagileak aukeratu.
"The doctor is choosing the referee."

1 2 3 4



(12) a. Sendagileak antzezlea aukeratu.
b. Antzezlea sendagileak aukeratu.
"The doctor is choosing the actor."

1 2 3 4



(13) a. Gidariak margolaria aukeratu.
b. Margolaria gidariak aukeratu.
"The pilot is choosing the painter."

1 2 3 4

(14) a. Gidariak sendagilea aukeratu.
b. Sendagilea gidariak aukeratu.
"The pilot is choosing the doctor."

1 2 3 4

(15) a. Gidariak epailea aukeratu.
b. Epailea gidariak aukeratu.
"The pilot is choosing the referee."

1 2 3 4

(16) a. Gidariak antzezlea aukeratu.
b. Antzezlea gidariak aukeratu.
"The pilot is choosing the actor."

1 2 3 4

(17) a. Margolariak sendagilea aukeratu.
b. Sendagilea margolariak aukeratu.
"The painter is choosing the doctor."

1 2 3 4

(18) a. Margolariak gidaria aukeratu.
b. Gidaria margolariak aukeratu.
"The painter is choosing the pilot."

1 2 3 4

(19) a. Margolariak epailea aukeratu.
b. Epailea margolariak aukeratu.
"The painter is choosing the referee."

1 2 3 4

(20) a. Margolariak antzezlea aukeratu.
b. Antzezlea margolariak aukeratu.
"The painter is choosing the actor."

1 2 3 4

(21) a. Antzezleak margolaria zelatatu.
b. Margolaria antzezleak zelatatu.
"The actor is spying on the painter."

1 2 3 4

(22) a. Antzezleak sendagilea zelatatu.
b. Sendagilea antzezleak zelatatu.
"The actor is spying on the doctor."

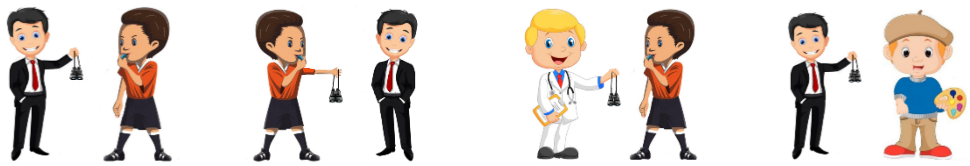
1 2 3 4

(23) a. Antzezleak gidaria zelatatu.
b. Gidaria antzezleak zelatatu.
"The actor is spying on the pilot."

1 2 3 4

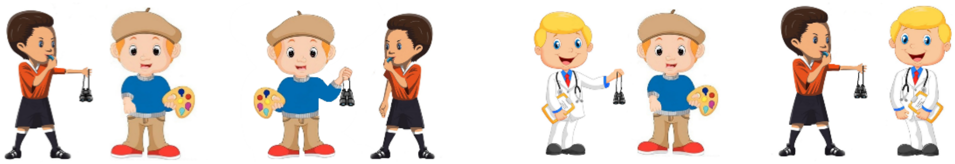
(24) a. Antzezleak epailea zelatatu.
b. Epailea antzezleak zelatatu.
"The actor is spying on the referee."

1 2 3 4



(25) a. Epaileak margolaria zelatatu.
b. Margolaria epaileak zelatatu.
"The referee is spying on the painter."

1 2 3 4



(26) a. Epaileak sendagilea zelatatu.
b. Sendagilea epaileak zelatatu.
"The referee is spying on the doctor."

1 2 3 4

(27) a. Epaileak gidaria zelatatu.
b. Gidaria epaileak zelatatu.
"The referee is spying on the pilot."

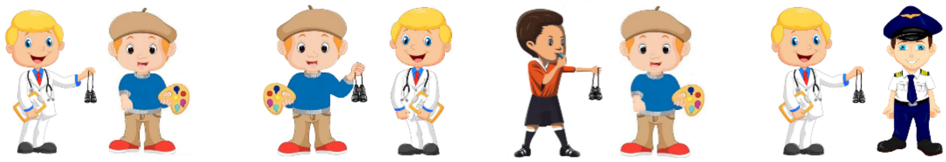
1 2 3 4

(28) a. Epaileak antzezlea zelatatu.
b. Antzezlea epaileak zelatatu.
"The referee is spying on the actor."

1 2 3 4

(29) a. Sendagileak margolaria zelatatu.
b. Margolaria sendagileak zelatatu.
"The doctor is spying on the painter."

1 2 3 4

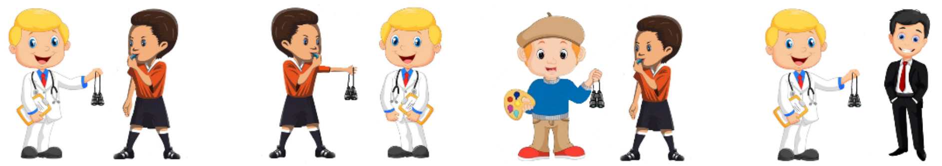


(30) a. Sendagileak gidaria zelatatu.
b. Gidaria sendagileak zelatatu.
"The doctor is spying on the pilot."

1 2 3 4

(31) a. Sendagileak epailea zelatatu.
b. Epailea sendagileak zelatatu.
"The doctor is spying on the referee."

1 2 3 4



(32) a. Sendagileak antzezlea zelatatu.
b. Antzezlea sendagileak zelatatu.
"The doctor is spying on the actor."

1 2 3 4

(33) a. Gidariak margolaria zelatatu.
b. Margolaria gidariak zelatatu.
"The pilot is spying on the painter."

1 2 3 4

(34) a. Gidariak sendagilea zelatatu.
b. Sendagilea gidariak zelatatu.
"The pilot is spying on the doctor."

1 2 3 4

(35) a. Gidariak epailea zelatatu.
b. Epailea gidariak zelatatu.
"The pilot is spying on the referee."

1 2 3 4

(36) a. Gidariak antzezlea zelatatu.
b. Antzezlea gidariak zelatatu.
"The pilot is spying on the actor."

1 2 3 4

(37) a. Margolariak sendagilea zelatatu.
b. Sendagilea margolariak zelatatu.
"The painter is spying on the doctor."

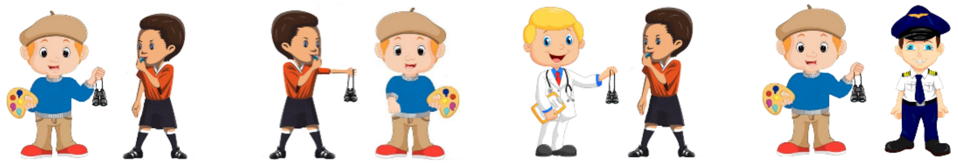
1 2 3 4

(38) a. Margolariak gidaria zelatatu.
b. Gidaria margolariak zelatatu.
"The painter is spying on the pilot."

1 2 3 4

(39) a. Margolariak epailea zelatatu.
b. Epailea margolariak zelatatu.
"The painter is spying on the referee."

1 2 3 4

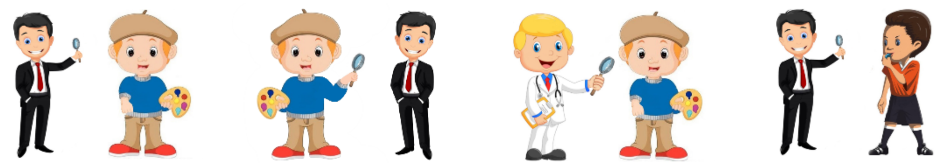


(40) a. Margolariak antzezlea zelatatu.
b. Antzezlea margolariak zelatatu.
"The painter is spying on the actor."

1 2 3 4

(41) a. Antzezleak margolaria aztertu.
b. Margolaria antzezleak aztertu.
"The actor is examining the painter."

1 2 3 4

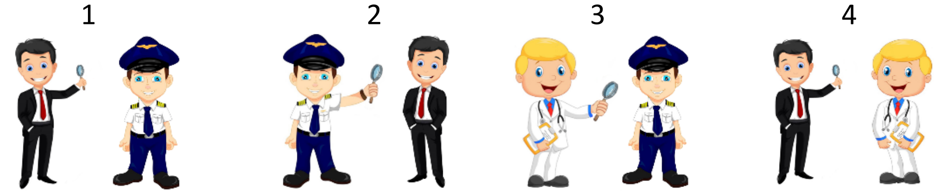


(42) a. Antzezleak sendagilea aztertu.
b. Sendagilea antzezleak aztertu.
"The actor is examining the doctor."

1 2 3 4

(43) a. Antzezleak gidaria aztertu.
b. Gidaria antzezleak aztertu.
"The actor is examining the pilot."

1 2 3 4

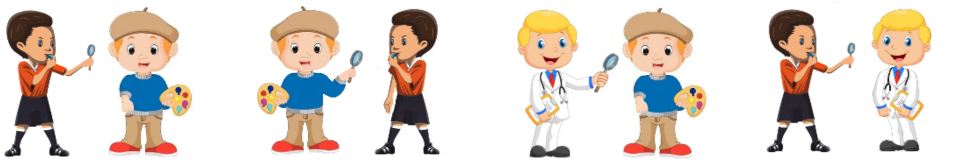


(44) a. Antzezleak epailea aztertu.
b. Epailea antzezleak aztertu.
"The actor is examining the referee."

1 2 3 4

(45) a. Epaileak margolaria aztertu.
b. Margolaria epaileak aztertu.
"The referee is examining the painter."

1 2 3 4

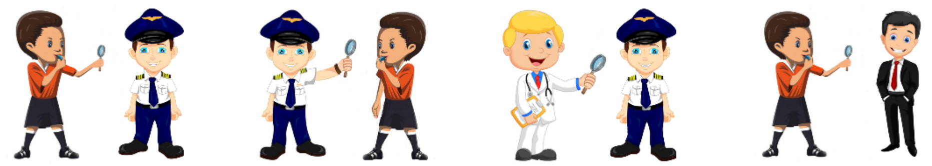


(46) a. Epaileak sendagilea aztertu.
b. Sendagilea epaileak aztertu.
"The referee is examining the doctor."

1 2 3 4

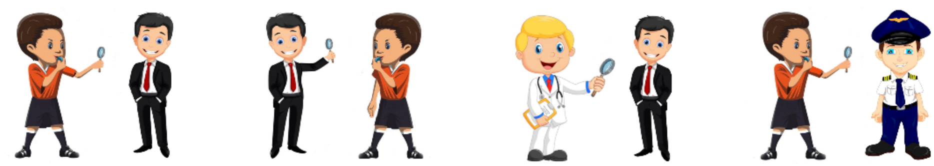
(47) a. Epaileak gidaria aztertu.
b. Gidaria epaileak aztertu.
"The referee is examining the pilot."

1 2 3 4



(48) a. Epaileak antzezlea aztertu.
b. Antzezlea epaileak aztertu.
"The referee is examining the actor."

1 2 3 4

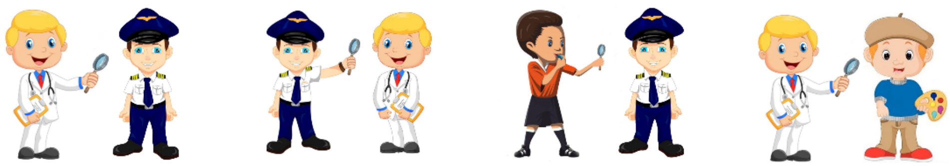


(49) a. Sendagileak margolaria aztertu.
b. Margolaria sendagileak aztertu.
"The doctor is examining the painter."

1 2 3 4

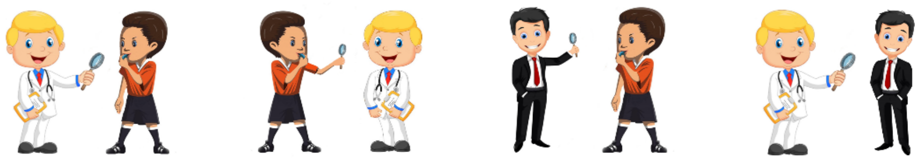
(50) a. Sendagileak gidaria aztertu.
b. Gidaria sendagileak aztertu.
"The doctor is examining the pilot."

1 2 3 4



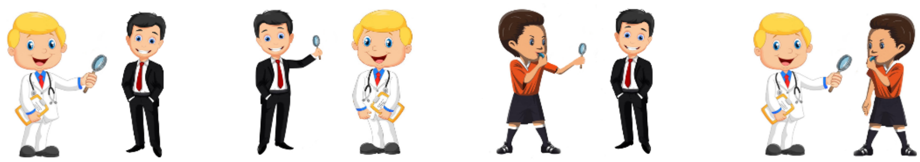
(51) a. Sendagileak epailea aztertu.
b. Epailea sendagileak aztertu.
"The doctor is examining the referee."

1 2 3 4



(52) a. Sendagileak antzezlea aztertu.
b. Antzezlea sendagileak aztertu.
"The doctor is examining the actor."

1 2 3 4



(53) a. Gidariak margolaria aztertu.
b. Margolaria gidariak aztertu.
"The pilot is examining the painter."

1 2 3 4

(54) a. Gidariak sendagilea aztertu.
b. Sendagilea gidariak aztertu.
"The pilot is examining the doctor."

1 2 3 4

(55) a. Gidariak epailea aztertu.
b. Epailea gidariak aztertu.
"The pilot is examining the referee."

1 2 3 4

(56) a. Gidariak antzezlea aztertu.
b. Antzezlea gidariak aztertu.
"The pilot is examining the actor."

1 2 3 4

(57) a. Margolariak sendagilea aztertu.
b. Sendagilea margolariak aztertu.
"The painter is examining the doctor."

1 2 3 4

(58) a. Margolariak gidaria aztertu.
b. Gidaria margolariak aztertu.
"The painter is examining the pilot."

1 2 3 4

(59) a. Margolariak epailea aztertu.
b. Epailea margolariak aztertu.
"The painter is examining the referee."

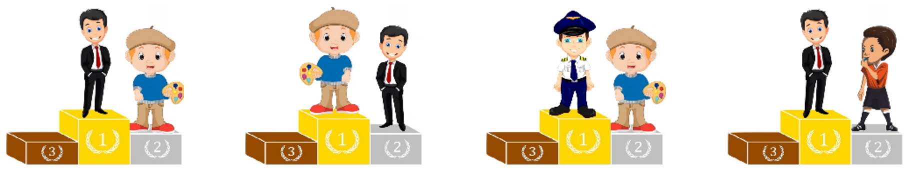
1 2 3 4

(60) a. Margolariak antzezlea aztertu.
b. Antzezlea margolariak aztertu.
"The painter is examining the actor."

1 2 3 4

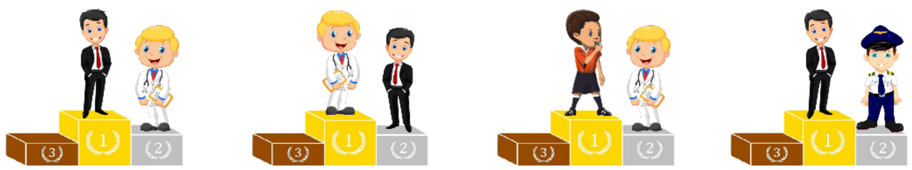
(61) a. Antzezleak margolaria gainditu.
b. Margolaria antzezleak gainditu.
"The actor is surpassing the painter."

1 2 3 4



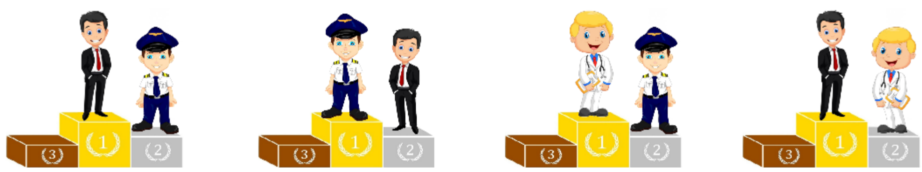
(62) a. Antzezleak sendagilea gainditu.
b. Sendagilea antzezleak gainditu.
"The actor is surpassing the doctor."

1 2 3 4



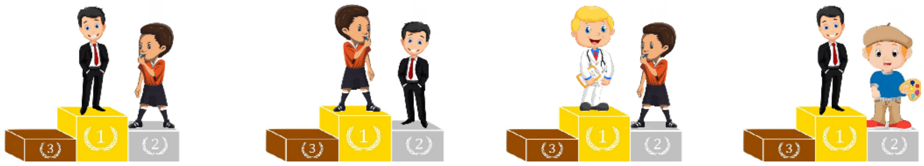
(63) a. Antzezleak gidaria gainditu.
b. Gidaria antzezleak gainditu.
"The actor is surpassing the pilot."

1 2 3 4



(64) a. Antzezleak epailea gainditu.
b. Epailea antzezleak gainditu.
"The actor is surpassing the referee."

1 2 3 4

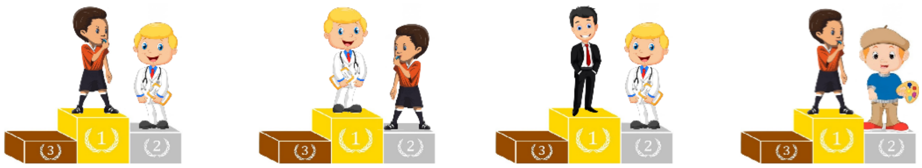


(65) a. Epaileak margolaria gainditu.
b. Margolaria epaileak gainditu.
"The referee is surpassing the painter."

1 2 3 4

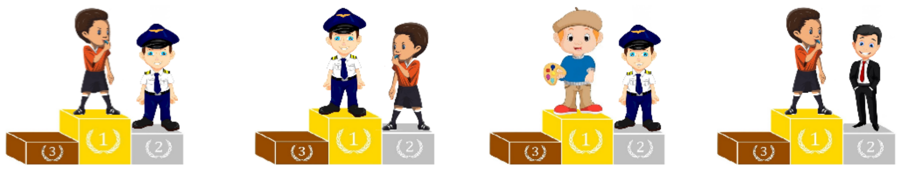
(66) a. Epaileak sendagilea gainditu.
b. Sendagilea epaileak gainditu.
"The referee is surpassing the doctor."

1 2 3 4



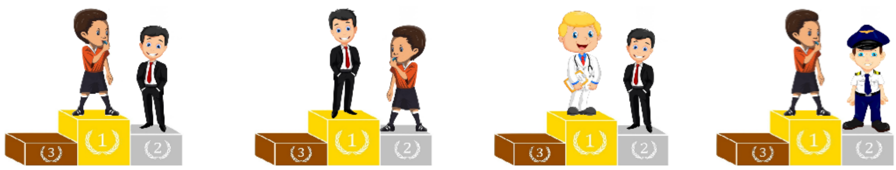
(67) a. Epaileak gidaria gainditu.
b. Gidaria epaileak gainditu.
"The referee is surpassing the pilot."

1 2 3 4



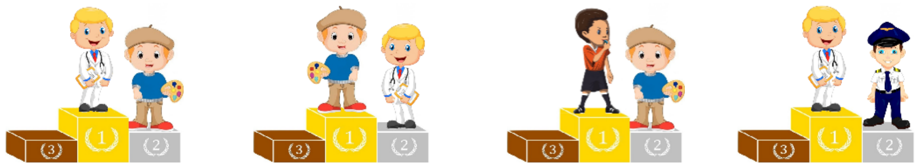
(68) a. Epaileak antzezlea gainditu.
b. Antzezlea epaileak gainditu.
"The referee is surpassing the actor."

1 2 3 4



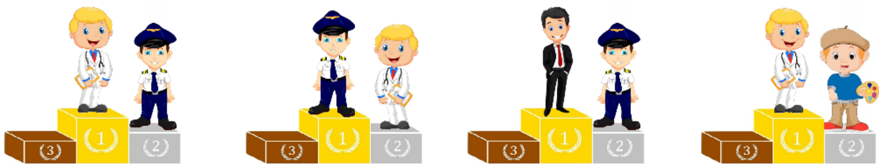
(69) a. Sendagileak margolaria gainditu.
b. Margolaria sendagileak gainditu.
"The doctor is surpassing the painter."

1 2 3 4



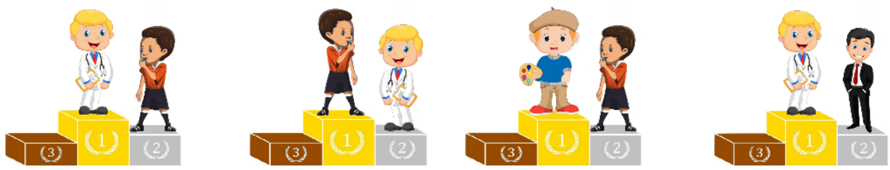
(70) a. Sendagileak gidaria gainditu.
b. Gidaria sendagileak gainditu.
"The doctor is surpassing the pilot."

1 2 3 4



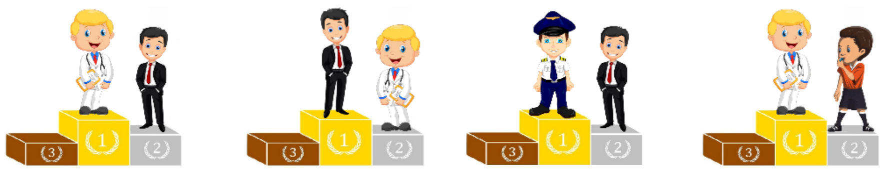
(71) a. Sendagileak epailea gainditu.
b. Epailea sendagileak gainditu.
"The doctor is surpassing the referee."

1 2 3 4



(72) a. Sendagileak antzezlea gainditu.
b. Antzezlea sendagileak gainditu.
"The doctor is surpassing the actor."

1 2 3 4



(73) a. Gidariak margolaria gainditu.
b. Margolaria gidariak gainditu.
"The pilot is surpassing the painter."

1 2 3 4

(74) a. Gidariak sendagilea gainditu.
b. Sendagilea gidariak gainditu.
"The pilot is surpassing the doctor."

1 2 3 4

(75) a. Gidariak epailea gainditu.
b. Epailea gidariak gainditu.
"The pilot is surpassing the referee."

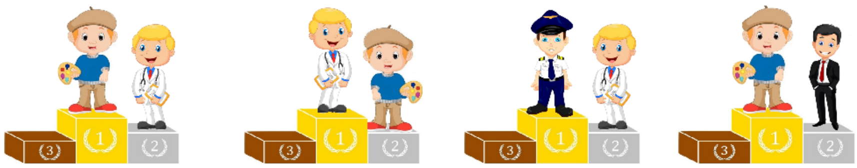
1 2 3 4

(76) a. Gidariak antzezlea gainditu.
b. Antzezlea gidariak gainditu.
"The pilot is surpassing the actor."

1 2 3 4

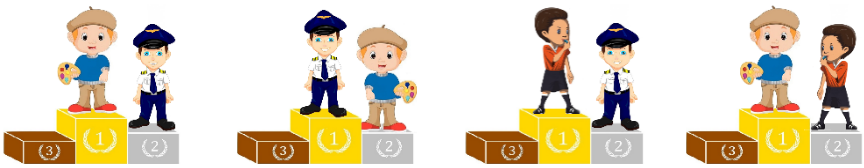
(77) a. Margolariak sendagilea gainditu.
b. Sendagilea margolariak gainditu.
"The painter is surpassing the doctor."

1 2 3 4



(78) a. Margolariak gidaria gainditu.
b. Gidaria margolariak gainditu.
"The painter is surpassing the pilot."

1 2 3 4



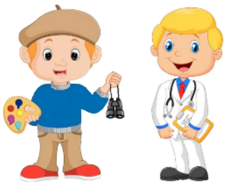
(79) a. Margolariak epailea gainditu.
b. Epailea margolariak gainditu.
"The painter is surpassing the referee."

1 2 3 4

(80) a. Margolariak antzezlea gainditu.
b. Antzezlea margolariak gainditu.
"The painter is surpassing the actor."

1 2 3 4

Testing set: Pictures used in the production task



6. Instructions used in Experiments 4 and 5

This appendix includes the instructions for the first vocabulary-learning phase, the exposure phase, the second vocabulary-learning phase, the testing phase, the debriefing phase and the reading span task in Experiments 4 and 5. Instructions are presented in Spanish, the language of the experiments. The English translation is presented below each slide.

First vocabulary-learning phase

¡Bienvenido/a!

En este experimento vas a aprender algunas palabras en euskera mediante imágenes. Aprenderás nombres y verbos.

Pulsa la BARRA DE ESPACIO para continuar.

ENG: Welcome! In this experiment you will learn some words in Basque with the help of pictures. You will learn nouns and verbs. Press the SPACE BAR to continue.

A continuación, verás una a una las imágenes que representan los nombres y los verbos. Debajo de cada imagen aparecerá escrito el nombre o el verbo que representa. Al mismo tiempo, oirás como suena. La traducción de cada palabra al castellano aparecerá entre paréntesis.

Tras escuchar cada palabra, deberás repetirla en voz alta. Para pasar de una imagen a la siguiente, pulsa la barra de espacio.

Tu tarea consiste en aprender estos nombres y verbos en euskera junto con las imágenes que describen. Fíjate bien en las imágenes y escucha cada palabra con atención.

¿Preparado/a?

Pulsa la BARRA DE ESPACIO para empezar.

ENG: You will now see the pictures representing the nouns and the verbs one by one. Below each picture it is written the noun or the verb it represents. At the same time, the word will be played. You will see the Spanish translation of each word between brackets. After listening to each word, you have to repeat it aloud. Press the space bar to go from one picture to the next one. Your task is to learn these nouns and verbs in Basque together with the pictures they describe. Pay close attention to the pictures and listen carefully to each word. Press the SPACE BAR to start.

Ahora realizarás un pequeño test para ver si has aprendido las palabras del euskera. Como en la anterior parte, verás las imágenes que representan los nombres y los verbos una a una, pero esta vez no te diremos qué palabra va asociada a cada imagen.

Debajo de cada imagen verás cuatro o cinco palabras y deberás escoger la opción correcta tan rápido como puedas.Para ello deberás usar los números 1, 2, 3, 4 (o 5) del teclado (1 para escoger la primera imagen, 2 para escoger la segunda imagen, etc).

Una vez escogida la palabra, verás un tick si has acertado y una cruz si has fallado. ¡Para poder pasar a la siguiente parte del experimento deberás acertar todas las palabras!

Pulsa la BARRA DE ESPACIO para empezar.

ENG: You will now take a little test to determine whether you have learnt the words. You will see the pictures that represent the nouns and verbs one by one, but this time you will not be told which word describes each picture. Below each picture, you will see four or five words and you will have to choose the correct option as fast as possible. To do this, you must press the keys 1, 2, 3, 4 or 5 (1 to select the first picture, 2 to select the second picture, and so on). Once you select a word, you will see a tick if your answer is correct and a cross if your answer is incorrect. To move on to the next part of the experiment you must make no mistakes! Press the SPACE BAR to start.

If the picture-word matching task has to be repeated:

Has cometido algún fallo.

¡Vuélvelo a intentar!

Pulsa la BARRA DE ESPACIO para empezar.

ENG: You have made at least one mistake. Try again! Press the SPACE BAR to start.

¡Bien hecho! A continuación vamos a ver si eres capaz de producir el nombre o verbo que describe cada imagen. Una vez más, verás las imágenes una a una. Sin embargo, no verás ni escucharás la palabra que las describe. Tu tarea consiste en pronunciar en voz alta la palabra asociada con cada imagen.
El ordenador va a enregistrar tus respuestas.

Cuando hayas pronunciado la palabra que crees que describe cada imagen, pulsa la barra de espacio para leer y escuchar la respuesta correcta.
Pulsa la barra de espacio para pasar de una imagen a la siguiente.

Pulsa la BARRA DE ESPACIO para empezar.

ENG: Well done! You will now be asked to name the noun or the verb that describes each picture. Once more, you will see the pictures one by one. Nevertheless, you will not hear nor see the word that describes each picture. Your task is to say aloud the word that describes each picture. Your answers will be recorded. After producing the word you think describes a picture, press the space bar to read and listen to the correct answer. Press the space bar to go from one picture to the next one. Press the SPACE BAR to start.

¡Muy bien! Es momento para un pequeño descanso.

Espera las instrucciones de la experimentadora.

ENG: Very good! Now it is time to take a short break. Wait for instructions from the experimenter.

<u>If the picture-naming task has to be repeated:</u>

Te has equivocado a la hora de producir algunas palabras. ¡Vuélvelo a intentar!

Una vez más, verás las imágenes una a una. Sin embargo, no verás ni escucharás la palabra que las describe.
Tu tarea consiste en pronunciar en voz alta la palabra asociada con cada imagen.
El ordenador va a enregistrar tus respuestas.

Cuando hayas pronunciado la palabra que crees que describe cada imagen, pulsa la barra de espacio para leer y escuchar la respuesta correcta.
Pulsa la barra de espacio para pasar de una imagen a la siguiente.

Pulsa la BARRA DE ESPACIO para empezar.

ENG: You have named at least one picture wrong. Try again! Once more, you will see pictures one by one. However, you will not see nor hear the words describing them. Your task is to name aloud the word associated with each picture. The computer will record your responses. Once you have named the word you think describes a picture, press the space bar to read and hear the correct answer. Press the space bar to move from one picture to the next one. Press the SPACE BAR to start.

Exposure phase

¡Bien hecho! Ahora te mostraremos los nombres y verbos que has aprendido en una oración. Verás imágenes una a una y podrás tanto leer como oir una oración describiendo cada imagen.

Tu tarea consiste en leer y escuchar la oración y observar la imagen correspondiente. Las imágenes irán pasando automáticamente, así que no es necesario que pulses ninguna tecla.

Pulsa la BARRA DE ESPACIO para empezar.

ENG: You will now see the nouns and verbs you learnt in a sentence. You will see pictures one by one, each accompanied by a sentence describing it. Your task is to look at the pictures and read and listen to the accompanying sentences. Pictures will be automatically presented one after the other, so you do not need to press any key. Press the SPACE BAR to start.

Second vocabulary-learning phase

A continuación aprenderás cuatro verbos nuevos. Aprenderás estos verbos de la misma forma que has aprendido el resto de palabras. Es decir:

Verás las imágenes que representan los verbos una a una. Debajo de cada imagen aparecerá el verbo escrito y la traducción al castellano entre paréntesis. Al mismo tiempo, oirás como suena cada verbo. Tras oir una palabra, deberás repetirla en voz alta.

Recuerda: deberás aprender los verbos y sus imágenes asociadas. Fíjate bien en las imágenes y escucha las palabras con atención. Pulsa la barra de espacio para pasar de un verbo al siguiente.

¿Preparado/a?

Pulsa la BARRA DE ESPACIO para empezar.

ENG: You will now learn four new verbs. You will learn them in the same way you have learnt the rest of the words. That is: You will see pictures representing the verbs one by one. Below each picture there will be a verb written and its Spanish translation between brackets. At the same time, the verb will be played. After listening to a verb, you must repeat it aloud. Remember: You must learn the verbs and the pictures associated with them. Have a close look at the pictures and listen carefully to the words. Press the space bar to go from one verb to the next one. Press the SPACE BAR to start.

¡Bien! Ya te has familiarizado con los cuatro verbos.
Ahora vamos a ver si has sido capaz de aprenderlos.

Verás las cuatro imágenes una a una. Debajo de cada imagen aparecerán escritos los cuatro verbos. Elige la opción correcta lo más rápido que puedas usando los números 1, 2, 3 o 4 del teclado.

Para pasar a la siguiente parte del experimento deberás acertar los cuatro verbos.

Pulsa la BARRA DE ESPACIO para empezar.

ENG: Good! You have familiarized yourself with the four verbs. Let's see if you have learnt them. You will see the four pictures one by one. The four verbs will be written below each picture. Select the correct option as fast as possible pressing the keys 1, 2, 3 or 4. To move on to the next part of the experiment, you must make no mistakes. Press the SPACE BAR to start.

<u>If the picture-word matching task has to be repeated:</u>

Has cometido algún fallo.

¡Vuélvelo a intentar!

Pulsa la BARRA DE ESPACIO para empezar.

ENG: You have made at least one mistake. Try again! Press the SPACE BAR to start.

¡Buen trabajo! Ahora es el turno de intentar pronunciar estos nuevos verbos en voz alta. Verás las imágenes una a una, pero no verás ni escucharás la palabra que las describe. Tu tarea consiste en producir la palabra asociada con cada imagen. Una vez más, el ordenador va a enregistrar tus respuestas.

Cuando hayas pronunciado la palabra que crees que describe cada imagen, pulsa la barra de espacio para leer y escuchar la respuesta correcta.
Pulsa la barra de espacio para pasar de una imagen a la siguiente.

Pulsa la BARRA DE ESPACIO para empezar.

ENG: Well done! Now it is time to pronounce the new verbs. You will see the pictures one by one, but you will not see nor hear the word describing them. You must name the word associated with each picture. Your answers will be recorded. After producing the word you think describes a picture, press the space bar to read and listen to the correct answer. Press the space bar to go from one picture to the next one. Press the SPACE BAR to start.

¡Bien! Es momento para otro pequeño descanso.

Espera las instrucciones de la experimentadora.

ENG: Good! Now it is time to take a short break. Wait for instructions from the experimenter.

If the picture-naming task has to be repeated:

Te has equivocado a la hora de producir algunas palabras. ¡Vuélvelo a intentar!

Una vez más, verás las imágenes una a una. Sin embargo, no verás ni escucharás la palabra que las describe.
Tu tarea consiste en pronunciar en voz alta la palabra asociada con cada imagen.
El ordenador va a enregistrar tus respuestas.

Cuando hayas pronunciado la palabra que crees que describe cada imagen, pulsa la barra de espacio para leer y escuchar la respuesta correcta.
Pulsa la barra de espacio para pasar de una imagen a la siguiente.

Pulsa la BARRA DE ESPACIO para empezar.

ENG: You have named at least one picture wrong. Try again! Once more, you will see pictures one by one. However, you will not see nor hear the words describing them. Your task is to name the word associated with each picture. The computer will record your responses. Once you have named the word you think describes a picture, press the space bar to read and hear the correct answer. Press the space bar to move from one picture to the next one. Press the SPACE BAR to start.

Testing phase

¡Genial! A continuación, verás y escucharás oraciones con los cinco nombres que ya sabes y los cuatro verbos que has aprendido hace un momento. Exceptuando estos verbos, las oraciones que escucharás serán IDÉNTICAS a las que has visto y escuchado antes.

Cada oración irá acompañada de una imagen. Tu tarea consiste en decidir, tan rápido como puedas, si la oración describe correctamente la imagen o no.

Pulsa A para indicar que la relación entre la imagen y la oración es CORRECTA.
Pulsa L para indicar que la relación entre la imagen y la oración es INCORRECTA.

Empezaremos con cuatro oraciones de práctica.

Pulsa la BARRA DE ESPACIO para empezar.

ENG: Great! You will now read and listen to sentences with the five nouns you already know and the four verbs you have just learnt. Except for these verbs, the sentences will be IDENTICAL to the ones you have read and heard before. Each sentence will be presented together with a picture. Your task is to decide, as quickly as possible, if the sentence is a correct description of the picture or not. Press A to indicate that the relationship between the picture and the sentence is CORRECT.

Press L to indicate that the relationship between the picture and the sentence is INCORRECT. We will start with four practice sentences. Press the SPACE BAR to start.

¡Bien! Ahora empezaremos con la prueba de verdad.

Escucharás unas cuantas oraciones, cada una acompañada de una imagen.
Tu tarea consiste en indicar tan rápido como puedas si la oración describe correctamente la imagen.

Recuerda:
Pulsa A para indicar que la relación entre la imagen y la oración es CORRECTA.
Pulsa L para indicar que la relación entre la imagen y la oración es INCORRECTA.

Pulsa la BARRA DE ESPACIO para empezar.

ENG: Good! You will now start with the test. You will read and listen to some sentences, each accompanied by a picture. You must indicate, as quickly as possible, if the sentence is a correct description of the picture or not. Remember: Press A to indicate that the relationship between the picture and the sentence is CORRECT. Press L to indicate that the relationship between the picture and the sentence is INCORRECT. Press the SPACE BAR to start.

Estás a punto de terminar el experimento.
Antes de irte, te pedimos que completes dos pequeñas tareas más.

A continuación verás 8 imágenes parecidas a las que has visto en el ejercicio anterior.Tu tarea consiste en escribir la oración que describe cada imagen.
Para ayudarte a recordar como se escriben los nombres y los verbos, estos aparecen a la derecha de la imagen.

Escribe la oración en el cuadro de texto que aparece debajo de la imagen.
Cuando termines, pulsa ENTER.
Si no sabes cual es la oración que describe una imagen, pulsa ENTER.

Pulsa la BARRA DE ESPACIO para empezar.

ENG: You are about to finish the experiment. Before you leave, we ask you to complete two more tasks. You will now see eight pictures. Your task is to write a sentence that describes each picture. To help you remember how the nouns and verbs are spelled, these appear to the right of the picture. Write the sentence in the text box that appears below the picture. When you are done, press ENTER. If you do not know which sentence describes a picture, press ENTER. Press the SPACE BAR to start.

Debriefing phase (Experiment 4)

En todas las oraciones que has escuchado había dos nombres (antzezle, gidari, sendagile, margolari o epaile).

¿Te has dado cuenta de que, cuando estos nombres aparecían en una oración, cambiaban respecto a su forma original?

¿Podrías decir qué cambio sufrían?

¿Sabrías decir por qué?

¿En que parte del experimento te has dado cuenta de este cambio en la forma de los nombres?

¿Has seguido alguna estrategia para realizar el test?

(Máximo 1.000 caracteres. El teclado no detecta acentos). Cuando termines, pulsa ENTER.

ENG: In all the sentences you have heard there were two nouns (*antzezle, gidari, sendagile, margolari* or *epaile*). Did you notice that when these nouns appeared in a sentence their original form changed? Could you say how did it change? Could you say why? In which part of the experiment did you notice the change in form? Did you follow any strategy to perform the test? (Maximum 1,000 characters. The keyboard does not admit accent marks). When you finish, press ENTER.

Debriefing phase (Experiment 5)

En todas las oraciones que has escuchado había dos nombres, un verbo y otra palabra.

¿Podrías decir cuál es esa palabra?

¿Sabrías decir por qué aparecía en la oración o cómo se usaba?

¿En qué parte del experimento te has dado cuenta de esto?

¿Has seguido alguna estrategia para realizar el test?

(Máximo 1.000 caracteres. El teclado no detecta acentos). Cuando termines, pulsa ENTER.

ENG: In all the sentences you have heard there were two nouns, a verb and another word. Could you say what this other word was? Could you say why it appeared in the sentence or how it was used? In which part of the experiment did you realize this? Did you follow any strategy to perform the test? (Maximum 1,000 characters. The keyboard does not admit accent marks). When you finish, press ENTER.

Final slide

> Has llegado al FINAL DEL EXPERIMENTO.
>
> ¡Muchas gracias por participar!

ENG: You have reached the END OF THE EXPERIMENT. Thank you for participating!

Reading span task

> En este experimento, tu tarea consiste en memorizar las letras que irán apareciendo en la pantalla a medida que leas las frases.
> A continuación, tendrás un breve entrenamiento para que veas cómo funciona este experimento.
>
> Haz clic para continuar leyendo las instrucciones.

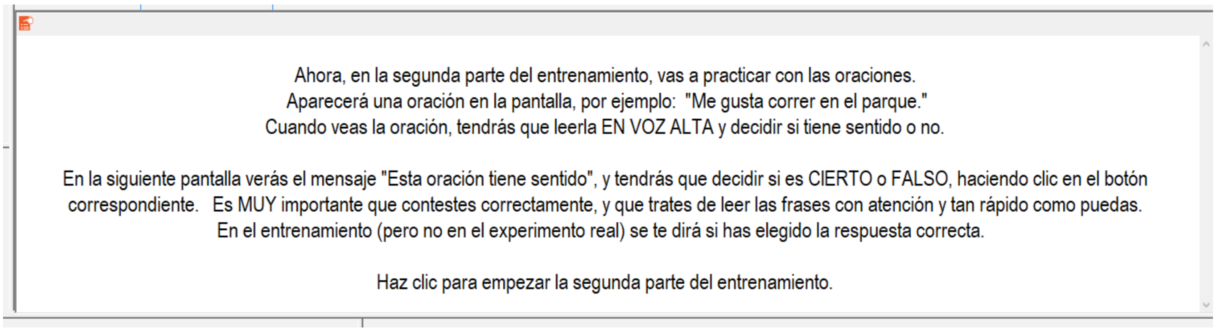
ENG: In this experiment, your task is to memorize the letters that will appear on the screen as you read sentences. Now, you will have a brief training so that you understand how the experiment works. Click to continue reading the instructions.

> La sesión de entrenamiento tiene tres partes. En esta primera parte, las letras irán apareciendo en la pantalla de una en una.
> Trata de recordarlas en el orden en el que hayan aparecido.
> Tras unas 2-3 letras, aparecerán12 letras en la pantalla, con un pequeño recuadro junto a cada una de ellas.
> Tu tarea consiste en seleccionar una por una las letras que hayas visto previamente, en el orden en el que hayan aparecido, marcando el recuadro correspondiente.
>
> Haz clic para continuar leyendo las instrucciones.

ENG: The training session has three parts. In the first part, letters will appear one by one on the screen. Try to remember them in the order in which they appear. After 2-3 letters, 12 letters will appear on the screen, with a small box next to each of them. Your task is to select one by one the letters you have previously seen, in the order in which they appeared, clicking on the corresponding box. Click to continue reading the instructions.

> Si cometes algún error, puedes hacer clic en BORRAR para empezar de nuevo.
>
> Recuerda que es muy importante que introduzcas las letras en el mismo orden en el que las hayas visto.
> Si olvidas alguna de las letras, haz clic en el recuadro EN BLANCO para indicar la posición de la letra que has olvidado.
>
> Haz clic para empezar la primera parte del entrenamiento.

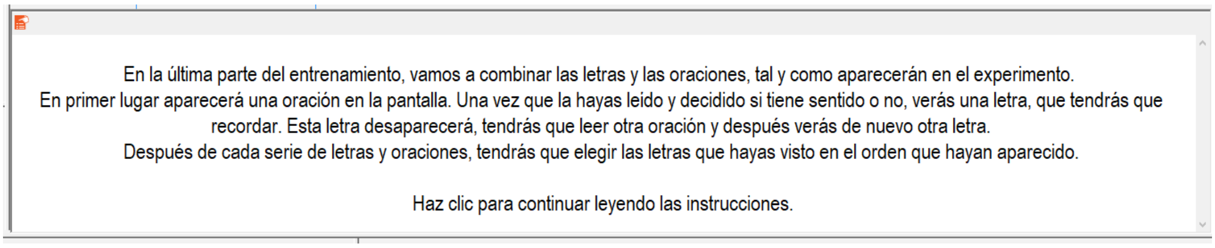
ENG: If you make a mistake, you can click on ERASE to start again. Remember that it is very important that you introduce the letters in the same order in which you have seen them. If you forget a letter, click on the box BLANK to indicate the position of the letter you forgot. Click to start the first part of the training.


Ahora, en la segunda parte del entrenamiento, vas a practicar con las oraciones.
Aparecerá una oración en la pantalla, por ejemplo: "Me gusta correr en el parque."
Cuando veas la oración, tendrás que leerla EN VOZ ALTA y decidir si tiene sentido o no.

En la siguiente pantalla verás el mensaje "Esta oración tiene sentido", y tendrás que decidir si es CIERTO o FALSO, haciendo clic en el botón correspondiente. Es MUY importante que contestes correctamente, y que trates de leer las frases con atención y tan rápido como puedas.
En el entrenamiento (pero no en el experimento real) se te dirá si has elegido la respuesta correcta.

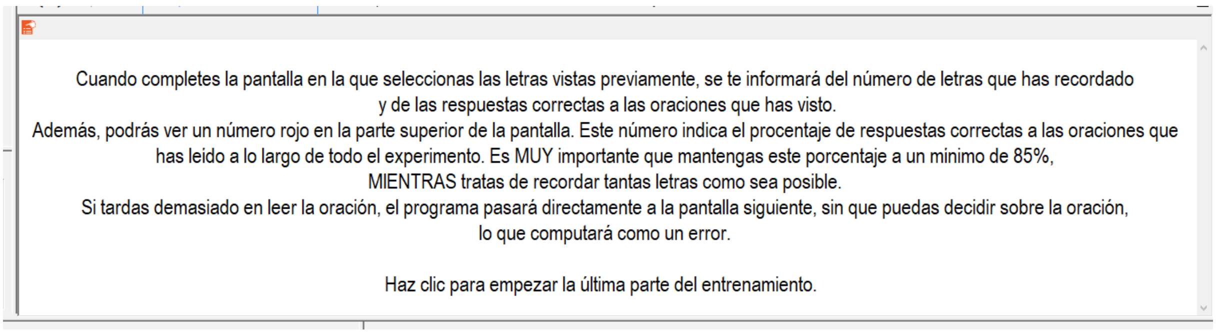
Haz clic para empezar la segunda parte del entrenamiento.

ENG: Now, in the second part of the training, you are going to practice with sentences. A sentence will appear on the screen, for example: "I like running in the park". When you see the sentence, you will have to read it ALOUD and decide if it makes sense or not. In the next screen you will see the message "This sentence makes sense" and you will have to decide if this is TRUE or FALSE, clicking on the corresponding option. It is VERY important that you answer correctly and that you read the sentences carefully and as quickly as you can. In the training (but not in the real experiment) you will be told if your answer was correct. Click to start the second part of the training.


En la última parte del entrenamiento, vamos a combinar las letras y las oraciones, tal y como aparecerán en el experimento.
En primer lugar aparecerá una oración en la pantalla. Una vez que la hayas leido y decidido si tiene sentido o no, verás una letra, que tendrás que recordar. Esta letra desaparecerá, tendrás que leer otra oración y después verás de nuevo otra letra.
Después de cada serie de letras y oraciones, tendrás que elegir las letras que hayas visto en el orden que hayan aparecido.

Haz clic para continuar leyendo las instrucciones.

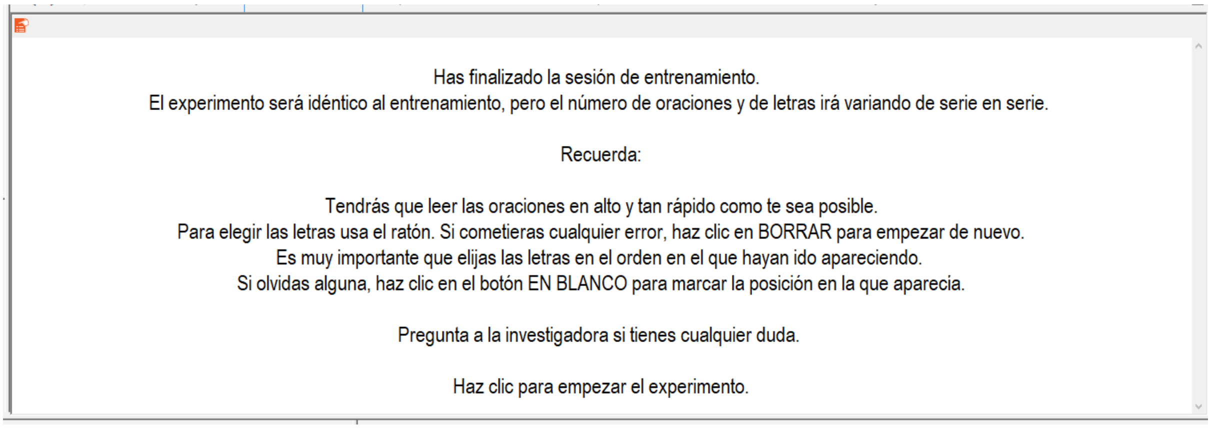
ENG: In the last part of the training, letters and sentences will be combined, just as they will appear in the experiment. First, a sentence will appear on the screen. Once you have read it and have decided if it makes sense or not, you will see a letter that you will have to remember. This letter will disappear, you will read another sentence and then you will see another letter. After each series of letters and sentences, you will have to select the letters you have seen in the order in which they appeared. Click to continue reading the instructions.



Cuando completes la pantalla en la que seleccionas las letras vistas previamente, se te informará del número de letras que has recordado
y de las respuestas correctas a las oraciones que has visto.
Además, podrás ver un número rojo en la parte superior de la pantalla. Este número indica el procentaje de respuestas correctas a las oraciones que
has leido a lo largo de todo el experimento. Es MUY importante que mantengas este porcentaje a un mínimo de 85%,
MIENTRAS tratas de recordar tantas letras como sea posible.
Si tardas demasiado en leer la oración, el programa pasará directamente a la pantalla siguiente, sin que puedas decidir sobre la oración,
lo que computará como un error.

Haz clic para empezar la última parte del entrenamiento.

ENG: Once you have completed the screen in which you select the letters you have previously seen, you will be told the number of letters you have correctly recalled and the number of correct responses to the sentences you have seen. Additionally, you will see a red number at the top of the screen. This number indicates the percentage of correct responses to the sentences that you have read throughout the experiment. It is VERY important that you keep that percentage to a minimum of 85%, while trying to remember as many letters as possible. If it takes you too long to read a sentence, the programme will automatically move on to the next screen, without letting you make a decision about the sentence. This will count as a mistake. Click to start the last part of the training.



Has finalizado la sesión de entrenamiento.
El experimento será idéntico al entrenamiento, pero el número de oraciones y de letras irá variando de serie en serie.

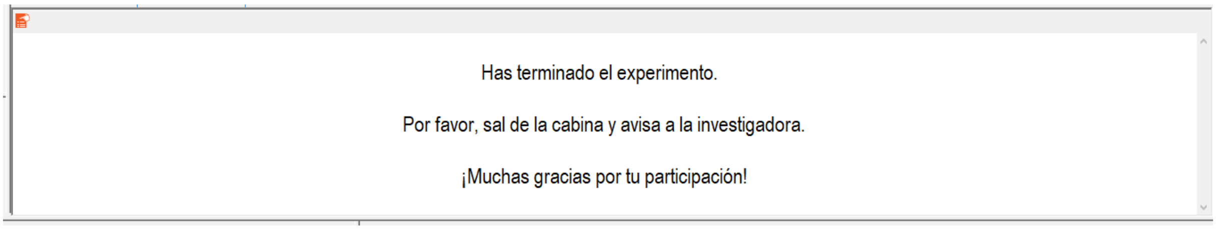
Recuerda:

Tendrás que leer las oraciones en alto y tan rápido como te sea posible.
Para elegir las letras usa el ratón. Si cometieras cualquier error, haz clic en BORRAR para empezar de nuevo.
Es muy importante que elijas las letras en el orden en el que hayan ido apareciendo.
Si olvidas alguna, haz clic en el botón EN BLANCO para marcar la posición en la que aparecia.

Pregunta a la investigadora si tienes cualquier duda.

Haz clic para empezar el experimento.

ENG: You have finished the training session. The experiment will be identical to the training, but the number of sentences and letters will vary from series to series. Remember: You will have to read the sentences aloud and as quickly as possible. Use the mouse to choose the letters. If you make a mistake, click on ERASE to start again. It is very important that you choose the letters in the order in which they appeared. If you forget a letter, click on BLANK to mark the position in which it appeared. Ask the researcher if you have any doubts. Click to start the experiment.



ENG: You have finished the experiment. Please, leave the booth and notify it to the researcher. Thank you for participating!

7. Rubric to evaluate awareness and transcription of verbal reports in Experiment 4

Rubric evaluating awareness

The questions asked in the verbal report were:

1. In all the sentences you have heard there were two nouns (*antzezle, gidari, sendagile, margolari* or *epaile*). Did you notice that when these nouns appeared in a sentence their original form changed?
2. Could you say how did it change?
3. Could you say why?
4. In which part of the experiment did you notice the change in form?
5. Did you follow any strategy to perform the test?

I present below the rubric used to evaluate awareness of agent-patient marking based on answers to questions 1-3.

Participant status	Description
Aware	The participant states that when a noun ended in *–ak*, it was the agent/subject of the sentence. Additionally or alternatively, the participant reports that when a noun ended in *–a*, it was the patient/object of the sentence.
Unaware	The participant is not able to identify the agent-patient marks. Alternatively, s/he is able to identify them but cannot (correctly) say what conceptual role they mark.

TABLE C-7.1. Rubric used to classify participants as *aware* or *unaware* of L2 agent-patient marking based on their responses in the verbal report of Experiment 4.

Verbal reports and awareness

In what follows, I present cognate and non-cognate learners' responses to the questions in the verbal report. The column "Report of agent-patient marking" shows answers to questions 1-3. The column "Awareness" indicates whether participants were classified as *aware* or *unaware* of agent-patient marking. Regarding question 4, all aware learners indicated that they noticed the change in the form of the nouns during the exposure phase. Finally, the column "Strategy test" indicates the strategy that each participant followed to perform the sentence-picture congruency task. Based on answers to question 5, I classified the strategies into three categories: (i) accept syntactic conditions and reject semantic conditions (i.e. accept syntactically congruent and incongruent sentence-picture pairs and reject semantically incongruent ones), (ii) apply agent-patient marking knowledge or (iii) intuition.

Cognate learners

Participant	Report of agent-patient marking	Awareness	Strategy test
P1	Sí, me he dado cuenta de que [los nombres] cambiaban.	Unaware	Accept syntactic conditions and reject semantic conditions
P2	El sujeto que hacía la acción acababa en –*ak*, el otro acababa en –*a*.	Aware	Apply agent-patient marking knowledge
P3	Sí, la persona que producía la acción acababa en –*ak* y la que la recibía, en –*a*.	Aware	Apply agent-patient marking knowledge
P4	La palabra que hacía de sujeto agregaba –*ak* al final y la que hacía de objeto directo agregaba una –*a*. Me he dado cuenta en la fase de aprender las oraciones que esas palabras estaban declinando.	Aware	Apply agent-patient marking knowledge
P5	A algunos nombres se les añadía la terminación –*ak*, al que está haciendo la acción.	Aware	Apply agent-patient marking knowledge
P6	Sí, se le añadía algo al final del nombre, supongo que son los pronombres.	Unaware	Accept syntactic conditions and reject semantic conditions
P7	A algunas palabras se les añaden terminaciones, ya sea la letra *k* o la *a*.	Unaware	Intuition
P8	Se escribe –*ak* al final de quien realiza la acción y –*a* al final de aquel que la recibe.	Aware	Accept syntactic conditions and reject semantic conditions
P9	La persona que realiza la acción lleva el sufijo –*eak*, mientras que la persona sobre la que recae la acción lleva otro sufijo, –*ea* o –*ia*.	Aware	Accept syntactic conditions and reject semantic conditions
P10	La oración se formulaba con dos nombres y un verbo. El nombre que terminaba en –*ak* era el que realizaba la acción y por ello se conjuga.	Aware	Intuition
P11	Sufijo –*k*. Este nombre era el sujeto del verbo.	Aware	Apply agent-patient marking knowledge
P12	Se añadía al final una –*a* o –*ak* en cada nombre. El por qué no me ha dado tiempo de entenderlo porque he visto que no era por el nombre en sí. Es decir, no por ser X nombre iba a llevar al final la –*a* o –*ak*.	Unaware	Intuition

Participant	Report of agent-patient marking	Awareness	Strategy test
P13	Al escribir las oraciones he intentado usar la sintaxis de sujeto, luego complemento directo y al final el verbo. Por ejemplo, si el sujeto era el pintor y el que recibe la acción era el árbitro, ponía primero *Margolari epaile* y al final el verbo. Si es *pintatu*, la oración final era *Margolari epaile pintatu.*	Unaware	Accept syntactic conditions and reject semantic conditions
P14	El sujeto que ejecutaba la acción llevaba detrás el sufijo *–eak*.	Aware	Apply agent-patient marking knowledge
P15	Sí que me he dado cuenta del cambio (*–ak*), he llegado a la conclusión de que es debido a la persona la cual hacía la acción.	Aware	Accept syntactic conditions and reject semantic conditions
P16	Se añaden sufijos a los nombres. A los nombres acabados en *–e* se les añade *–ak* y a los que acaban en *–i*, *–a*. Supongo que se hace así para que haya concordancia.	Unaware	Intuition
P17	Se incluía una *–k* al final. Creo que probablemente la *–k* final servía para indicar el sujeto de la oración.	Aware	Apply agent-patient marking knowledge
P18	No he apreciado que los nombres cambiaban respecto a su forma original cuando aparecían en una oración. Me he basado en mirar primeramente si el verbo que ponía correspondía con la imagen y después si los dos nombres que aparecían correspondían con los dibujos también.	Unaware	Accept syntactic conditions and reject semantic conditions
P19	Sí, en algunas palabras se añadía una *–a* al final y en otras una *–a* y una *–k*, dependiendo del que hacía la acción. He intuido que el que hacía la acción era el que llevaba la *–k*.	Aware	Intuition
P20	No, no me he dado cuenta de que las formas sufriesen ningún cambio.	Unaware	Intuition
P21	Creo que se les añade una *–k* final.	Unaware	Accept syntactic conditions and reject semantic conditions
P22	Cambiaba la forma del nombre que recibía la acción añadiendo el sufijo *–k* o *–ek* dependiendo de si terminaba en vocal o consonante.	Unaware	Accept syntactic conditions and reject semantic conditions

Participant	Report of agent-patient marking	Awareness	Strategy test
P23	Las palabras que cambiaban su forma se referían a la persona que realizaba la acción, añadiendo *–ak* al final de la palabra.	Aware	Apply agent-patient marking knowledge
P24	Me he fijado en que los nombres que aparecían [en la imagen] concordasen con la persona que hacía la acción y que la persona que hacía la acción fuera seguida del verbo.	Unaware	Accept syntactic conditions and reject semantic conditions
P25	Sí, cuando realizaban la acción se le añadía *–ak* al final del nombre.	Aware	Apply agent-patient marking knowledge
P26	Al nombre de las personas que hacían una acción se le ponía un *–ak* al final de la palabra y se situaba al principio de la oración. El verbo iba siempre al final de la oración.	Aware	Accept syntactic conditions and reject semantic conditions
P27	Cuando la persona era la que realizaba la acción, al nombre de su profesión se le añadía una *–ak* al final.	Aware	Accept syntactic conditions and reject semantic conditions
P28	En una oración, quien hacía la acción tenía una *–k* al final de la palabra, y quien la recibía no sufría ningún cambio en la palabra.	Aware	Apply agent-patient marking knowledge
P29	El sujeto que realizaba la acción acababa con la terminación *–ak* y el que recibía esta acción acababa con *–a*.	Aware	Apply agent-patient marking knowledge
P30	Diferencia en los nombres añadiendo *–ak* o *–a* al final de cada nombre. Posiblemente se relaciona con quien hace la acción y quien la recibe, respectivamente. También cambiaba el orden de los nombres. El verbo no cambiaba, siempre estaba al final.	Aware	Apply agent-patient marking knowledge

TABLE C-7.2. Transcription of cognate learners' responses in the verbal report of Experiment 4 (in Spanish). Based on these responses, participants are classified as *aware* or *unaware* of L2 agent-patient marking. Each participant's reported strategy in the sentence-picture congruency task is also indicated.

English translation of verbal reports

P1: Yes, I noticed that [the nouns] changed.

P2: The subject doing the action ended in *–ak*, the other [noun] ended in *–a*.

P3: Yes, the person who did the action ended in *–ak* and the person who received it ended in *–a*.

P4: The word that acted as the subject had *–ak* at the end and the word that acted as the direct object had an *–a*. I noticed that words were declining in the phase of learning the sentences.

P5: The ending *–ak* was added to some nouns, to the ones doing the action.

P6: Yes, something was added to the end of each noun, I guess it is a pronoun.

P7: Endings are added to some words, either the letter *k* or *a*.

P8: *–ak* is written at the end of the person who performs the action and *–a,* at the end of the person who receives it.

P9: The person who performs the action has the suffix *–eak*, while the person on whom the action is performed has another suffix, *–ea* or *–ia*.

P10: Each sentence was constructed with two nouns and a verb. The noun that ended in *–ak* was the one that performed the action and this is why it was conjugated.

P11: Suffix *–k*. This noun was the subject of the verb.

P12: An *–a* or *–ak* was added at the end of each noun. I have not had time to understand why, because I saw that it was not because of the noun itself. That is, not because it was X noun, was it going to have *–a* or *–ak* at the end.

P13: When writing the sentences I have tried to use the syntax of subject, then direct object and the verb at the end. For example, if the subject was the painter and the person receiving the action was the referee, I wrote *Margolari epaile* first and the verb last. If this was *pintatu*, the final sentence was *Margolari epaile pintatu*.

P14: The subject who carried out the action had the suffix *–eak*.

P15: Yes, I have noticed the change (*–ak*), I have come to the conclusion that it marks the person who does the action.

P16: Suffixes are added to nouns.*–Ak* is added to nouns ending in *–e* and *–a,* to nouns ending in *–i*. I suppose it is done this way so that there is agreement.

P17: A *–k* was included at the end. I think that the final *–k* probably indicated the subject of the sentence.

P18: I have not noticed that the nouns changed from their original form when they appeared in a sentence. I based my answers on whether the verb and nouns in the sentence coincided with the ones in the picture or not.

P19: Yes, an *–a* was added at the end of some words, and an *–a* and a *–k* at the end of others, depending on who was doing the action. The one doing the action was the one with the *–k*.

P20: No, I have not noticed that the form [of nouns] changed.

P21: I think that a final *–k* was added.

P22: The form of the noun that received the action changed by adding the suffix *–k* or *–ek* depending on whether it ended in a vowel or a consonant.

P23: Words changing their form referred to the person who performed the action, *–ak* was added to the end of the word.

P24: I focused on whether the nouns that appeared [in the picture] matched the person doing the action and whether the person doing the action was followed by the verb.

P25: Yes, *–ak* was added at the end of the noun performing the action.

P26: The noun of the person who performed the action was added an *–ak* at the end and it was placed at the beginning of the sentence. The verb always came at the end of the sentence.
P27: An *–ak* was added to the profession of the person who performed the action.
P28: In a sentence, whoever did the action had a *–k* at the end of the word, and whoever received it did not undergo any change in the word.
P29: The subject who performed the action ended in *–ak* and the one who received this action ended in *–a*.
P30: Adding *–ak* or *–a* at the end of each noun. Possibly, it is related to who does the action and who receives it, respectively. The order of the nouns also changed. The verb did not change, it was always at the end.

Non-cognate learners

Participant	**Report of agent-patient marking**	**Awareness**	**Strategy test**
P1	Al final de cada nombre había un *–eak* o *–ak*. [Una de estas dos marcas] puede ser una conjugación para indicar quién hace la acción.	Unaware	Intuition
P2	Sí, me he dado cuenta de que estos nombres cambiaban respecto a su forma original, ya que no me cuadraban con lo que me había estudiado; aparecían más largos, pero no sé por qué. Interpreto que será por una cuestión gramatical o de relación con el verbo de la oración.	Unaware	Accept syntactic conditions and reject semantic conditions
P3	Cambio en la terminación. Errores en la descripción de la imagen con el texto, no coinciden.	Unaware	Intuition
P4	El cambio que sufrían las palabras se corresponde con la adhesión del sufijo *–leak* y *–a*. Creo que se corresponde con el hecho de quién realiza o sufre la acción, si es complemento directo o indirecto, pero no sé qué sufijo va con cada función.	Unaware	Intuition
P5	Si las tres palabras de la oración coincidían con la imagen, la daba por correcta. Si alguna no coincidía, la daba por incorrecta.	Unaware	Accept syntactic conditions and reject semantic conditions
P6	Sí, se les añadía un *–ea* o *–eak*, creo recordar. Al que se le añade *–eak* es el que realiza la acción.	Aware	Apply agent-patient marking knowledge
P7	Había sufijo *–ak* cuando el nombre era sujeto y sufijo *–a* cuando era complemento indirecto.	Aware	Apply agent-patient marking knowledge

Participant	**Report of agent-patient marking**	**Awareness**	**Strategy test**
P8	Me he dado cuenta de que cambia la palabra porque se le pone un sufijo, pero no sabría decir cuál ni por qué. En el ejercicio de correcto-incorrecto me he fijado más en los personajes. Si el personaje no aparecía [en la imagen], le daba a incorrecto y si los dos aparecían, me fijaba en el verbo. No me he fijado en el orden de las oraciones y en el ejercicio donde ponían las frases creo que a veces con una misma imagen cambiaba el orden de los personajes, no sé cuál va primero.	Unaware	Accept syntactic conditions and reject semantic conditions
P9	Me he dado cuenta de que [los nombres] se modificaban, creo que se modificaba el que hacía dicha acción. No sabría decir qué cambio sufren y me he dado cuenta de esto cuando aparecieron las frases con las imágenes rápido.	Unaware	Intuition
P10	Sí, imagino que cambiaban según su función en la frase. Si era sujeto tenía una terminación y si era complemento directo tenía otra. Algunas de las terminaciones que creo recordar son *–urtzu* y *–aile*.	Unaware	Intuition
P11	Me he dado cuenta que en el ejercicio en el que salían las oraciones automáticamente, los personajes que hacían las acciones tenían la terminación *–iak*. He intentado aplicar esto en el ejercicio posterior en el que se tenía que indicar si las oraciones son correctas o no.	Aware	Apply agent-patient marking knowledge
P12	Los nombres tenían dos terminaciones distintas que supongo que están relacionadas con quién hace y recibe la acción.	Unaware	Intuition
P13	Creo que el sujeto de la oración pasa a acabar en *–k* mientras que quien recibe la acción, el complemento directo, solo acaba en *–a*.	Aware	Apply agent-patient marking knowledge
P14	Creo que se añadía una *–a* a algunas palabras y otras acababan en *–ak*, pero no sé por qué.	Unaware	Accept syntactic conditions and reject semantic conditions
P15	Los nombres terminan en *–ak* cuando se refieren al sujeto que hace la acción.	Aware	Apply agent-patient marking knowledge

Participant	Report of agent-patient marking	Awareness	Strategy test
P16	He visto que los sustantivos acababan en *–ak* o *–a*. Me he dado cuenta en la fase del experimento en la que veía y oía las frases. Creo que la persona que acababa en *–ak*, como *margolariak*, era el receptor de la acción y la persona que acababa en *–a* era el sujeto o el que hace la acción.	Unaware	Apply agent-patient marking knowledge
P17	Algunas de las oraciones acaban en *–ak*. En ocasiones salían los dos nombres seguidos de un verbo y en otras nombre-verbo-nombre. No sabría decir por qué.	Unaware	Accept syntactic conditions and reject semantic conditions
P18	Me fijaba en quién hacía la acción, que era la primera palabra, y si coincidía [con la imagen] y estaba bien, luego miraba la segunda palabra, y si también estaba bien el oficio, por último miraba el verbo.	Unaware	Accept syntactic conditions and reject semantic conditions
P19	Me he dado cuenta de que a la persona que realiza la acción se le pone el sufijo *–ak* detrás. El orden de los nombres de los personajes iba cambiando.	Aware	Accept syntactic conditions and reject semantic conditions
P20	Me he dado cuenta que en las diferentes palabras había varias terminaciones que se añadían, como la *–a* y *–ak*. Creo que estas terminaciones sirven para determinar quién hace la acción y quién la recibe, pero no sé cuál es cuál.	Unaware	Accept syntactic conditions and reject semantic conditions
P21	Sí, las terminaciones de algunas de las palabras cambiaban, los nombres acababan en *–k* y podía ser en quien recaía la acción o bien quien la llevaba a cabo, no me ha dado tiempo a fijarme bien.	Unaware	Intuition
P22	Sí, uno de los dos nombres terminaba con la letra *k*.	Unaware	Accept syntactic conditions and reject semantic conditions
P23	Sí, se le sumaban letras al final. Creo que es porque dependiendo de si era el que daba la acción o la recibía, tenía una especie de conjugación y posición en la oración.	Unaware	Accept syntactic conditions and reject semantic conditions
P24	Sí, porque, por ejemplo, la persona que realizaba la acción, su nombre se escribía con una vocal y una *–k* al final.	Aware	Apply agent-patient marking knowledge
P25	Sí, los que realizaban la acción cambiaban su terminación por *–ak*.	Aware	Apply agent-patient marking knowledge

Participant	Report of agent-patient marking	Awareness	Strategy test
P26	Las palabras sufrían un cambio al final y los verbos continuaban igual. Sé que había dos cambios al final de los nombres, pero solo recuerdo uno, el *–ak*. Supongo que los cambios a final de palabra determinan quién hace y recibe la acción. He pensado que el nombre que hace la acción es el que termina en *–ak* y que aparece en la primera posición de la oración.	Aware[1]	Apply agent-patient marking knowledge
P27	Se agrega una terminación *–ak* al sujeto que realiza la acción del verbo hacia el que la recibe.	Aware	Accept syntactic conditions and reject semantic conditions
P28	He inferido que uno era el sujeto y el otro el complemento indirecto. El sujeto ganaba el sufijo *–oak* o similar, el otro nombre también tenía una terminación diferente, en *–a* o en *–e*.	Aware	Apply agent-patient marking knowledge
P29	Se le ponía una *–a* al final, posiblemente por la influencia de la siguiente palabra.	Unaware	Intuition
P30	Las palabras cambiaban cuando esos nombres eran quienes hacían la acción. Lo que ocurría era que se modificaban y se añadía una *–k* al final.	Aware	Apply agent-patient marking knowledge

TABLE C-7.3. Transcription of non-cognate learners' responses in the verbal report of Experiment 4 (in Spanish). Based on these responses, participants are classified as *aware* or *unaware* of L2 agent-patient marking. Each participant's reported strategy in the sentence-picture congruency task is also indicated.

English translation of verbal reports

P1: At the end of each noun there was an *–eak* or *–ak*. [One of these two marks] can be a conjugation to indicate who does the action.

P2: Yes, I have realized that nouns changed from their original form, since they did not fit with what I had studied; they appeared longer, but I do not know why. I interpret that it is due to a grammatical issue or a relationship with the verb of the sentence.

P3: Change in termination. Errors in the description of the picture with the text, they do not match.

P4: The change that the words underwent corresponds to the addition of the suffixes *–leak* and *–a*. I think this depends on who performs or receives the action, whether it is a direct or indirect object, but I do not know which suffix marks each function.

[1] This participant is considered *aware* of agent-patient marking because s/he correctly reports that *–ak* is attached to the subject/agent of the sentence. Yet, s/he incorrectly assumes that all sentences are subject-initial (see also participant 26 within cognate learners).

P5: If the three words in the sentence matched the picture, I considered it correct. If some did not match, I considered it incorrect.

P6: Yes, an *–ea* or *–eak* was added. The noun to which *–eak* is added is the one that performs the action.

P7: Adding the suffix *–ak* when the noun was a subject and the suffix *–a* when it was an indirect object.

P8: I have noticed that the words change because a suffix is added, but I could not say which one or why. In the correct-incorrect exercise, I paid attention to the characters. If a character did not appear [in the picture], I considered it incorrect. If both appeared, I looked at the verb. I have not paid attention to the word order in the sentences. In the exercise in which sentences were played, I think that sometimes the order of the characters changed, I do not know which one goes first.

P9: I have noticed that [the nouns] were modified. I think the person who performed the action was modified. I could not say what change it underwent. I realized this [in the part of the experiment] in which the sentences and pictures appeared fast.

P10: Yes, I imagine they changed depending on their function in the sentence. If it was a subject, it had one ending and if it was a direct object, it had another ending. Some of the endings that I remember are *–urtzu* and *–aile*.

P11: I noticed that in the exercise in which the sentences were presented automatically, the character who performed the action had the ending *–iak*. I tried to apply this in the subsequent exercise in which I had to indicate whether the sentences were correct or not.

P12: The nouns had two different endings that I assume are related to who does and receives the action.

P13: I think that the subject of the sentence ends in *–k* while the person receiving the action, the direct object, ends in *–a*.

P14: I think an *–a* was added to some words and others ended in *–ak*, but I do not know why.

P15: The nouns end in *–ak* when referring to the subject doing the action.

P16: I have seen that nouns ended in *–ak* or *–a*. I realized this in the phase of the experiment in which I saw and heard the sentences. I believe that the person ending in *–ak*, like *margolariak*, was the recipient of the action and the person ending in *–a* was the subject or the doer of the action.

P17: Some of the sentences end in *–ak*. Sometimes the two nouns appeared followed by a verb and sometimes the order was noun-verb-noun. I cannot say why.

P18: I looked at who was doing the action, which was the first word. If it coincided [with the picture], I looked at the second word. If it also coincided with the picture, I finally looked at the verb.

P19: I have noticed that the person who performs the action has the suffix *–ak*. The order of the characters changed.

P20: I have noticed that several endings were added to words, such as *–a* and *–ak*. I think these endings determine who does the action and who receives it, but I do not know which is which.

P21: Yes, the endings of some of the words changed. The noun ending in *–k* could be who received the action or who carried it out, I did not have time to look at it closely.

P22: Yes, one of the two nouns ended with the letter *k*.

P23: Yes, letters were added at the end. I think that depending on whether the noun was doing the action or receiving it, it had a specific conjugation and position in the sentence.

P24: Yes. For example, the noun of the person who performed the action was written with a vowel and a *–k* at the end.

P25: Yes, the ending of nouns performing the action changed to –*ak*.

P26: The words underwent a change at the end and the verbs remained the same. I know there were two changes at the end of the nouns, but I only remember one, the –*ak*. I suppose that the changes at the end of the words determine who does and receives the action. I think that the noun that does the action is the one that ends in –*ak* and it appears in the first position of the sentence.

P27: An –*ak* is added to the subject that performs the action of the verb.

P28: I have inferred that a noun was the subject and the other, the indirect object. The subject had the suffix –*oak* or similar. The other noun also had a different ending, –*a* or –*e*.

P29: An –*a* was written at the end, possibly due to the influence of the following word.

P30: The words changed when they were the ones doing the action. What happened was that a –*k* was added at the end.

8. Experimental materials used in Experiment 5

<u>Exposure set: Sentence-picture pairs</u>

Sentence-picture pairs constituting the exposure set in Experiment 5. Sentence-picture pairs are presented in groups of four, corresponding to the two conditions in the exposure phase in the two versions of the mini-language: a. SVO sentence-picture pair (cognate version), b. OVS sentence-picture pair (cognate version), c. SVO sentence-picture pair (non-cognate version) and d. OVS sentence-picture pair (non-cognate version). The English translation for each group of sentences is the same as for the equivalent SOV and OSV sentences in Experiment 4.

(1) a. Antzezle kastigatu a margolari.
b. A margolari kastigatu antzezle.
c. Antzezle zigortu a margolari.
d. A margolari zigortu antzezle.

(2) a. Antzezle kastigatu a sendagile.
b. A sendagile kastigatu antzezle.
c. Antzezle zigortu a sendagile.
d. A sendagile zigortu antzezle.

(3) a. Antzezle kastigatu a gidari.
b. A gidari kastigatu antzezle.
c. Antzezle zigortu a gidari.
d. A gidari zigortu antzezle.

(4) a. Antzezle kastigatu a epaile.
b. A epaile kastigatu antzezle.
c. Antzezle zigortu a epaile.
d. A epaile zigortu antzezle.

(5) a. Epaile kastigatu a margolari.
b. A margolari kastigatu epaile.
c. Epaile zigortu a margolari.
d. A margolari zigortu epaile.

(6) a. Epaile kastigatu a sendagile.
b. A sendagile kastigatu epaile.
c. Epaile zigortu a sendagile.
d. A sendagile zigortu epaile.

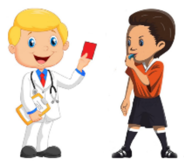
(7) a. Epaile kastigatu a gidari.
b. A gidari kastigatu epaile.
c. Epaile zigortu a gidari.
d. A gidari zigortu epaile.

(8) a. Epaile kastigatu a antzezle.
b. A antzezle kastigatu epaile.
c. Epaile zigortu a antzezle.
d. A antzezle zigortu epaile.

(9) a. Sendagile kastigatu a margolari.
b. A margolari kastigatu sendagile.
c. Sendagile zigortu a margolari.
d. A margolari zigortu sendagile.

(10) a. Sendagile kastigatu a gidari.
b. A gidari kastigatu sendagile.
c. Sendagile zigortu a gidari.
d. A gidari zigortu sendagile.

(11) a. Sendagile kastigatu a epaile.
b. A epaile kastigatu sendagile.
c. Sendagile zigortu a epaile.
d. A epaile zigortu sendagile.



(12) a. Sendagile kastigatu a antzezle.
b. A antzezle kastigatu sendagile.
c. Sendagile zigortu a antzezle.
d. A antzezle zigortu sendagile.

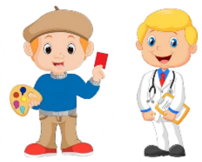
(13) a. Gidari kastigatu a margolari.
b. A margolari kastigatu gidari.
c. Gidari zigortu a margolari.
d. A margolari zigortu gidari.

(14) a. Gidari kastigatu a sendagile.
b. A sendagile kastigatu gidari.
c. Gidari zigortu a sendagile.
d. A sendagile zigortu gidari.

(15) a. Gidari kastigatu a epaile.
b. A epaile kastigatu gidari.
c. Gidari zigortu a epaile.
d. A epaile zigortu gidari.

(16) a. Gidari kastigatu a antzezle.
b. A antzezle kastigatu gidari.
c. Gidari zigortu a antzezle.
d. A antzezle zigortu gidari.

(17) a. Margolari kastigatu a sendagile.
b. A sendagile kastigatu margolari.
c. Margolari zigortu a sendagile.
d. A sendagile zigortu margolari.



(18) a. Margolari kastigatu a gidari.
b. A gidari kastigatu margolari.
c. Margolari zigortu a gidari.
d. A gidari zigortu margolari.

(19) a. Margolari kastigatu a epaile.
b. A epaile kastigatu margolari.
c. Margolari zigortu a epaile.
d. A epaile zigortu margolari.

(20) a. Margolari kastigatu a antzezle.
b. A antzezle kastigatu margolari.
c. Margolari zigortu a antzezle.
d. A antzezle zigortu margolari.

(21) a. Antzezle pintatu a margolari.
b. A margolari pintatu antzezle.
c. Antzezle margotu a margolari.
d. A margolari margotu antzezle.

(22) a. Antzezle pintatu a sendagile.
b. A sendagile pintatu antzezle.
c. Antzezle margotu a sendagile.
d. A sendagile margotu antzezle.

(23)
a. Antzezle pintatu a gidari.
b. A gidari pintatu antzezle.
c. Antzezle margotu a gidari.
d. A gidari margotu antzezle.

(24)
a. Antzezle pintatu a epaile.
b. A epaile pintatu antzezle.
c. Antzezle margotu a epaile.
d. A epaile margotu antzezle.

(25)
a. Epaile pintatu a margolari.
b. A margolari pintatu epaile.
c. Epaile margotu a margolari.
d. A margolari margotu epaile.

(26)
a. Epaile pintatu a sendagile.
b. A sendagile pintatu epaile.
c. Epaile margotu a sendagile.
d. A sendagile margotu epaile.

(27)
a. Epaile pintatu a gidari.
b. A gidari pintatu epaile.
c. Epaile margotu a gidari.
d. A gidari margotu epaile.

(28)
a. Epaile pintatu a anztezle.
b. A antzezle pintatu epaile.
c. Epaile margotu a antzezle.
d. A antzezle margotu epaile.

(29)
a. Sendagile pintatu a margolari.
b. A margolari pintatu sendagile.
c. Sendagile margotu a margolari.
d. A margolari margotu sendagile.

(30)
a. Sendagile pintatu a gidari.
b. A gidari pintatu sendagile.
c. Sendagile margotu a gidari.
d. A gidari margotu sendagile.

(31)
a. Sendagile pintatu a epaile.
b. A epaile pintatu sendagile.
c. Sendagile margotu a epaile.
d. A epaile margotu sendagile.

(32)
a. Sendagile pintatu a antzezle.
b. A antzezle pintatu sendagile.
c. Sendagile margotu a antzezle.
d. A antzezle margotu sendagile.

(33)
a. Gidari pintatu a margolari.
b. A margolari pintatu gidari.
c. Gidari margotu a margolari.
d. A margolari margotu gidari.

(34)
a. Gidari pintatu a sendagile.
b. A sendagile pintatu gidari.
c. Gidari margotu a sendagile.
d. A sendagile margotu gidari.

(35)
a. Gidari pintatu a epaile.
b. A epaile pintatu gidari.
c. Gidari margotu a epaile.
d. A epaile margotu gidari.

(36)
a. Gidari pintatu a antzezle.
b. A antzezle pintatu gidari.
c. Gidari margotu a anztezle.
d. A antzezle margotu gidari.

(37)
a. Margolari pintatu a sendagile.
b. A sendagile pintatu margolari.
c. Margolari margotu a sendagile.
d. A sendagile margotu margolari.

(38)
a. Margolari pintatu a gidari.
b. A gidari pintatu margolari.
c. Margolari margotu a gidari.
d. A gidari margotu margolari.

(39)
a. Margolari pintatu a epaile.
b. A epaile pintatu margolari.
c. Margolari margotu a epaile.
d. A epaile margotu margolari.

(40)
a. Margolari pintatu a anztezle.
b. A antzezle pintatu margolari.
c. Margolari margotu a antzezle.
d. A antzezle margotu margolari.

(41)
a. Antzezle presentatu a margolari.
b. A margolari presentatu antzezle.
c. Antzezle aurkeztu a margolari.
d. A margolari aurkeztu antzezle.

(42)
a. Antzezle presentatu a sendagile.
b. A sendagile presentatu antzezle.
c. Antzezle aurkeztu a sendagile.
d. A sendagile aurkeztu antzezle.

(43)
a. Antzezle presentatu a gidari.
b. A gidari presentatu antzezle.
c. Antzezle aurkeztu a gidari.
d. A gidari aurkeztu anztezle.

(44)
a. Antzezle presentatu a epaile.
b. A epaile presentatu antzezle.
c. Antzezle aurkeztu a epaile.
d. A epaile aurkeztu antzezle.

(45)
a. Epaile presentatu a margolari.
b. A margolari presentatu epaile.
c. Epaile aurkeztu a margolari.
d. A margolari aurkeztu epaile.

(46)
a. Epaile presentatu a sendagile.
b. A sendagile presentatu epaile.
c. Epaile aurkeztu a sendagile.
d. A sendagile aurkeztu epaile.

(47) a. Epaile presentatu a gidari.
b. A gidari presentatu epaile.
c. Epaile aurkeztu a gidari.
d. A gidari aurkeztu epaile.

(48) a. Epaile presentatu a anztezle.
b. A antzezle presentatu epaile.
c. Epaile aurkeztu a antzezle.
d. A antzezle aurkeztu epaile.

(49) a. Sendagile presentatu a margolari.
b. A margolari presentatu sendagile.
c. Sendagile aurkeztu a margolari.
d. A margolari aurkeztu sendagile.

(50) a. Sendagile presentatu a gidari.
b. A gidari presentatu sendagile.
c. Sendagile aurkeztu a gidari.
d. A gidari aurkeztu sendagile.

(51) a. Sendagile presentatu a epaile.
b. A epaile presentatu sendagile.
c. Sendagile aurkeztu a epaile.
d. A epaile aurkeztu sendagile.

(52) a. Sendagile presentatu a antzezle.
b. A antzezle presentatu sendagile.
c. Sendagile aurkeztu a antzezle.
d. A antzezle aurkeztu sendagile.

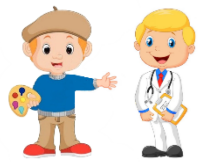
(53) a. Gidari presentatu a margolari.
b. A margolari presentatu gidari.
c. Gidari aurkeztu a margolari.
d. A margolari aurkeztu gidari.

(54) a. Gidari presentatu a sendagile.
b. A sendagile presentatu gidari.
c. Gidari aurkeztu a sendagile.
d. A sendagile aurkeztu gidari.

(55) a. Gidari presentatu a epaile.
b. A epaile presentatu gidari.
c. Gidari aurkeztu a epaile.
d. A epaile aurkeztu gidari.

(56) a. Gidari presentatu a anztezle.
b. A antzezle presentatu gidari.
c. Gidari aurkeztu a anztezle.
d. A antzezle aurkeztu gidari.

(57) a. Margolari presentatu a sendagile.
b. A sendagile presentatu margolari.
c. Margolari aurkeztu a sendagile.
d. A sendagile aurkeztu margolari.



(58) a. Margolari presentatu a gidari.
b. A gidari presentatu margolari.
c. Margolari aurkeztu a gidari.
d. A gidari aurkeztu margolari.

(59) a. Margolari presentatu a epaile.
b. A epaile presentatu margolari.
c. Margolari aurkeztu a epaile.
d. A epaile aurkeztu margolari.

(60) a. Margolari presentatu a antzezle.
b. A antzezle presentatu margolari.
c. Margolari aurkeztu a antzezle.
d. A antzezle aurkeztu margolari.

(61) a. Antzezle salutatu a margolari.
b. A margolari salutatu antzezle.
c. Antzezle agurtu a margolari.
d. A margolari agurtu antzezle.

(62) a. Antzezle salutatu a sendagile.
b. A sendagile salutatu anztezle.
c. Antzezle agurtu a sendagile.
d. A sendagile agurtu antzezle.

(63) a. Antzezle salutatu a gidari.
b. A gidari salutatu antzezle.
c. Antzezle agurtu a gidari.
d. A gidari agurtu antzezle.

(64) a. Antzezle salutatu a epaile.
b. A epaile salutatu antzezle.
c. Antzezle agurtu a epaile.
d. A epaile agurtu antzezle.

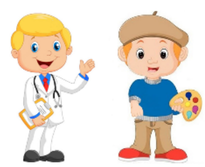
(65) a. Epaile salutatu a margolari.
b. A margolari salutatu epaile.
c. Epaile agurtu a margolari.
d. A margolari agurtu epaile.

(66) a. Epaile salutatu a sendagile.
b. A sendagile salutatu epaile.
c. Epaile agurtu a sendagile.
d. A sendagile agurtu epaile.

(67) a. Epaile salutatu a gidari.
b. A gidari salutatu epaile.
c. Epaile agurtu a gidari.
d. A gidari agurtu epaile.

(68) a. Epaile salutatu a antzezle.
b. A antzezle salutatu epaile.
c. Epaile agurtu a antzezle.
d. A antzezle agurtu epaile.

(69) a. Sendagile salutatu a margolari.
b. A margolari salutatu sendagile.
c. Sendagile agurtu a margolari.
d. A margolari agurtu sendagile.



(70) a. Sendagile salutatu a gidari.
b. A gidari salutatu sendagile.
c. Sendagile agurtu a gidari.
d. A gidari agurtu sendagile.

(71)
a. Sendagile salutatu a epaile.
b. A epaile salutatu sendagile.
c. Sendagile agurtu a epaile.
d. A epaile agurtu sendagile.

(72)
a. Sendagile salutatu a antzezle.
b. A antzezle salutatu sendagile.
c. Sendagile agurtu a antzezle.
d. A antzezle agurtu sendagile.

(73)
a. Gidari salutatu a margolari.
b. A margolari salutatu gidari.
c. Gidari agurtu a margolari.
d. A margolari agurtu gidari.

(74)
a. Gidari salutatu a sendagile.
b. A sendagile salutatu gidari.
c. Gidari agurtu a sendagile.
d. A sendagile agurtu gidari.

(75)
a. Gidari salutatu a epaile.
b. A epaile salutatu gidari.
c. Gidari agurtu a epaile.
d. A epaile agurtu gidari.

(76)
a. Gidari salutatu a antzezle.
b. A antzezle salutatu gidari.
c. Gidari agurtu a antzezle.
d. A antzezle agurtu gidari.

(77)
a. Margolari salutatu a sendagile.
b. A sendagile salutatu margolari.
c. Margolari agurtu a sendagile.
d. A sendagile agurtu margolari.

(78)
a. Margolari salutatu a gidari.
b. A gidari salutatu margolari.
c. Margolari agurtu a gidari.
d. A gidari agurtu margolari.

(79) a. Margolari salutatu a epaile.
b. A epaile salutatu margolari.
c. Margolari agurtu a epaile.
d. A epaile agurtu margolari.

(80) a. Margolari salutatu a antzezle.
b. A antzezle salutatu margolari.
c. Margolari agurtu a antzezle.
d. A antzezle agurtu margolari.

Testing set: Sentence-picture pairs used in the sentence-picture congruency task

Sentence-picture pairs constituting the testing set in Experiment 5. Sentences are presented in pairs: a. SVO sentence and b. OVS sentence. Each pair of sentences is associated with four pictures, resulting in the four experimental conditions in the sentence-picture congruency task: a/b-1. Syntactically congruent sentence-picture pair, a/b-2. Syntactically incongruent sentence-picture pair, a/b-3. Semantically incongruent sentence-picture pair with agent violation and a/b-4. Semantically incongruent sentence-picture pair with patient violation. The English translation for each group of sentences is the same as for the equivalent SOV and OSV sentences in Experiment 4.

(1) a. Antzezle aukeratu a margolari.
b. A margolari aukeratu antzezle.

1

2

3

4


(2) a. Antzezle aukeratu a sendagile.
b. A sendagile aukeratu antzezle.

1

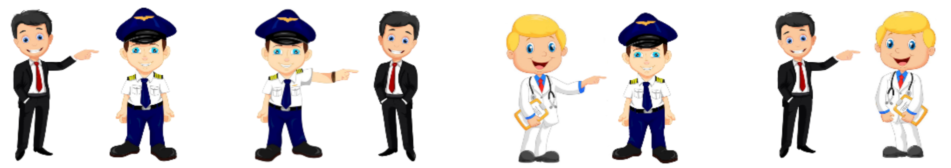
2

3

4

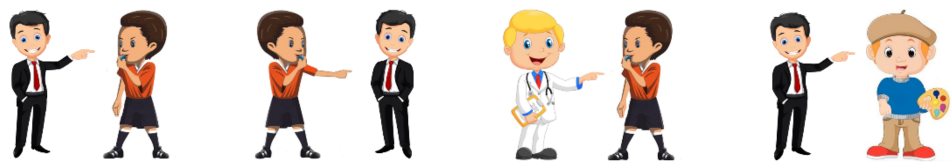

(3) a. Antzezle aukeratu a gidari.
b. A gidari aukeratu antzezle.

1 2 3 4



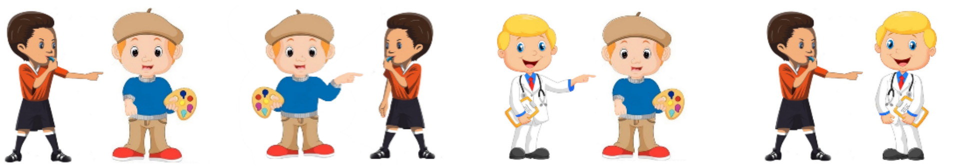
(4) a. Antzezle aukeratu a epaile.
b. A epaile aukeratu antzezle.

1 2 3 4



(5) a. Epaile aukeratu a margolari.
b. A margolari aukeratu epaile.

1 2 3 4

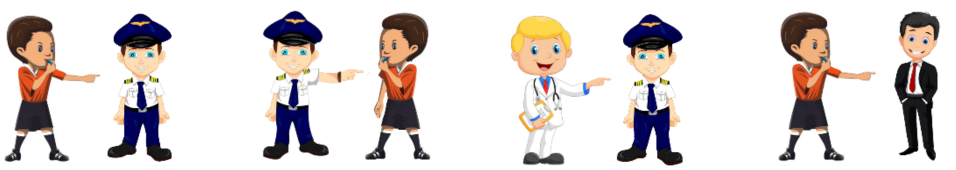


(6) a. Epaile aukeratu a sendagile.
b. A sendagile aukeratu epaile.

1 2 3 4

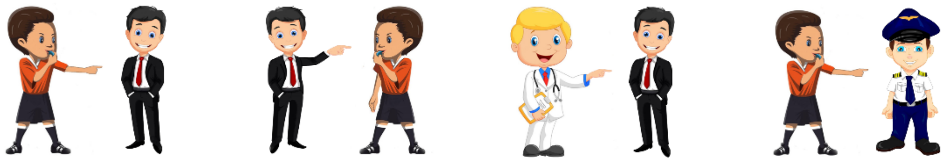
(7) a. Epaile aukeratu a gidari.
b. A gidari aukeratu epaile.

1 2 3 4



(8) a. Epaile aukeratu a antzezle.
b. A antzezle aukeratu epaile.

1 2 3 4

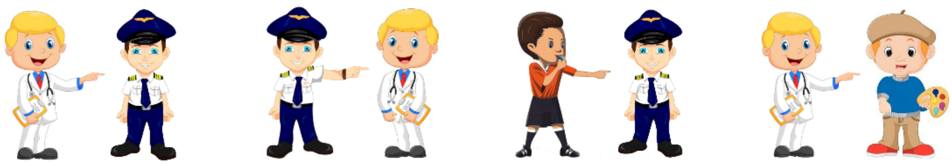


(9) a. Sendagile aukeratu a margolari.
b. A margolari aukeratu sendagile.

1 2 3 4

(10) a. Sendagile aukeratu a gidari.
b. A gidari aukeratu sendagile.

1 2 3 4

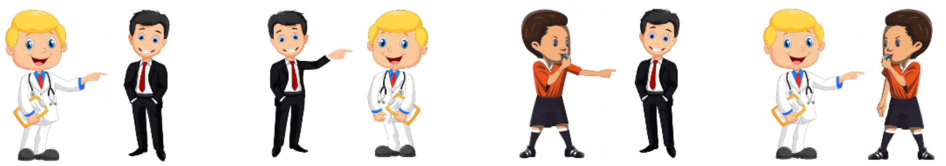


(11) a. Sendagile aukeratu a epaile.
b. A epaile aukeratu sendagile.

1 2 3 4

(12) a. Sendagile aukeratu a antzezle.
b. A antzezle aukeratu sendagile.

1 2 3 4



(13) a. Gidari aukeratu a margolari.
b. A margolari aukeratu gidari.

1 2 3 4

(14) a. Gidari aukeratu a sendagile.
b. A sendagile aukeratu gidari.

1 2 3 4

(15) a. Gidari aukeratu a epaile.
b. A epaile aukeratu gidari.

1 2 3 4

(16) a. Gidari aukeratu a antzezle.
b. A antzezle aukeratu gidari.

1 2 3 4

(17) a. Margolari aukeratu a sendagile.
b. A sendagile aukeratu margolari.

1 2 3 4

(18) a. Margolari aukeratu a gidari.
b. A gidari aukeratu margolari.

1 2 3 4

(19) a. Margolari aukeratu a epaile.
b. A epaile aukeratu margolari.

1 2 3 4

(20) a. Margolari aukeratu a antzezle.
b. A antzezle aukeratu margolari.

1 2 3 4

(21) a. Antzezle zelatatu a margolari.
b. A margolari zelatatu antzezle.

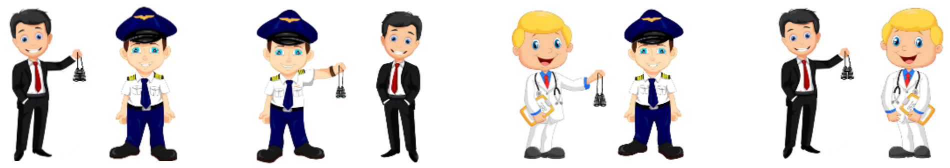
1 2 3 4

(22) a. Antzezle zelatatu a sendagile.
b. A sendagile zelatatu antzezle.

1 2 3 4

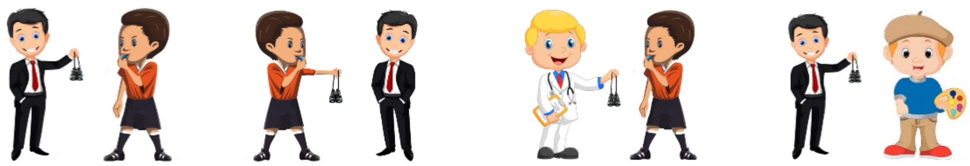
(23) a. Antzezle zelatatu a gidari.
b. A gidari zelatatu antzezle.

1 2 3 4



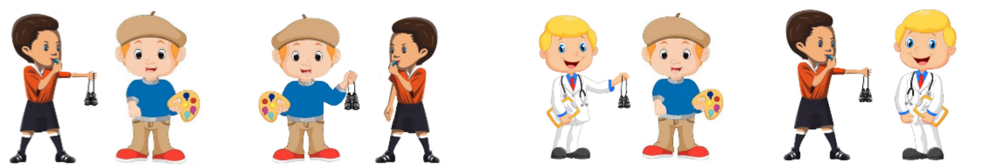
(24) a. Antzezle zelatatu a epaile.
b. A epaile zelatatu antzezle.

1 2 3 4



(25) a. Epaile zelatatu a margolari.
b. A margolari zelatatu epaile.

1 2 3 4



(26) a. Epaile zelatatu a sendagile.
b. A sendagile zelatatu epaile.

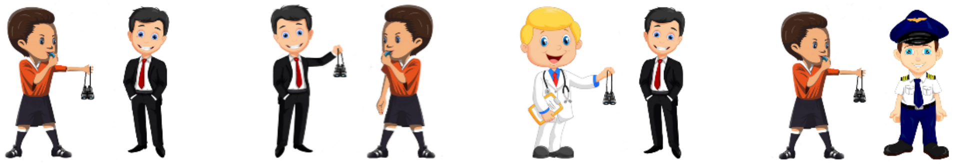
1 2 3 4

(27) a. Epaile zelatatu a gidari.
b. A gidari zelatatu epaile.

1 2 3 4

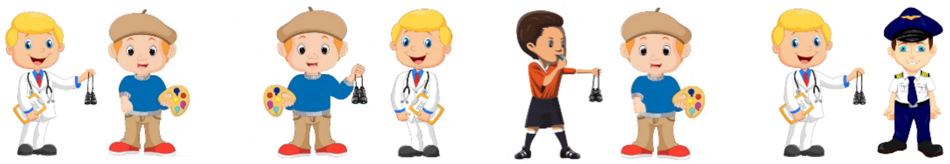
(28) a. Epaile zelatatu a antzezle.
b. A antzezle zelatatu epaile.

1 2 3 4



(29) a. Sendagile zelatatu a margolari.
b. A margolari zelatatu sendagile.

1 2 3 4

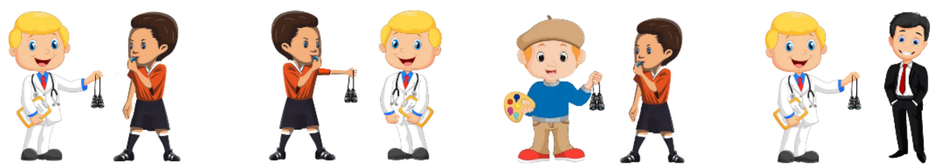


(30) a. Sendagile zelatatu a gidari.
b. A gidari zelatatu sendagile.

1 2 3 4

(31) a. Sendagile zelatatu a epaile.
b. A epaile zelatatu sendagile.

1 2 3 4



(32) a. Sendagile zelatatu a antzezle.
b. A antzezle zelatatu sendagile.

1 2 3 4

(33) a. Gidari zelatatu a margolari.
b. A margolari zelatatu gidari.

1 2 3 4

(34) a. Gidari zelatatu a sendagile.
b. A sendagile zelatatu gidari.

1 2 3 4

(35) a. Gidari zelatatu a epaile.
b. A epaile zelatatu gidari.

1 2 3 4

(36) a. Gidari zelatatu a antzezle.
b. A antzezle zelatatu gidari.

1 2 3 4

(37) a. Margolari zelatatu a sendagile.
b. A sendagile zelatatu margolari.

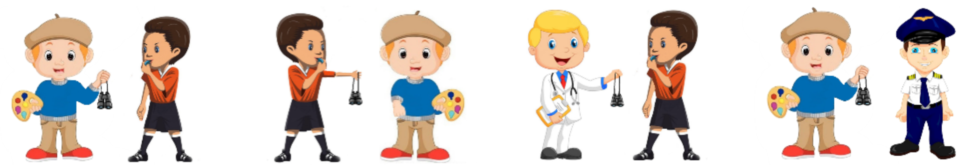
1 2 3 4

(38) a. Margolari zelatatu a gidari.
b. A gidari zelatatu margolari.

1 2 3 4

(39) a. Margolari zelatatu a epaile.
b. A epaile zelatatu margolari.

1 2 3 4



(40) a. Margolari zelatatu a antzezle.
b. A antzezle zelatatu margolari.

1 2 3 4

(41) a. Antzezle aztertu a margolari.
b. A margolari aztertu antzezle.

1 2 3 4

(42) a. Antzezle aztertu a sendagile.
b. A sendagile aztertu antzezle.

1 2 3 4

(43) a. Antzezle aztertu a gidari.
b. A gidari aztertu antzezle.

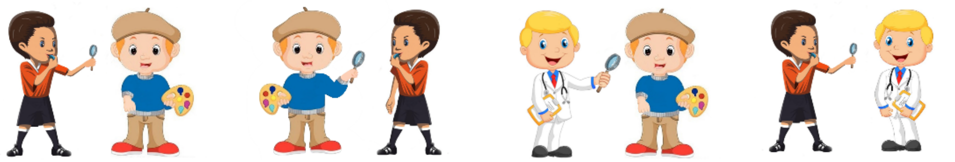
1 2 3 4

(44) a. Antzezle aztertu a epaile.
b. A epaile aztertu antzezle.

1 2 3 4

(45) a. Epaile aztertu a margolari.
b. A margolari aztertu epaile.

1 2 3 4

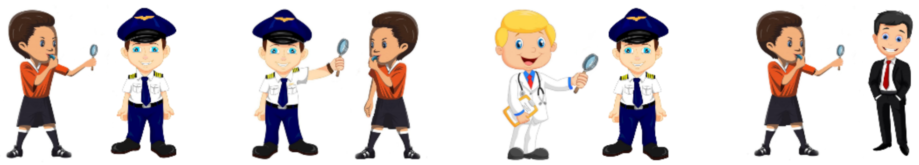


(46) a. Epaile aztertu a sendagile.
b. A sendagile aztertu epaile.

1 2 3 4

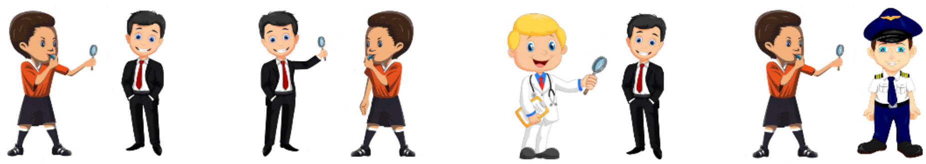
(47) a. Epaile aztertu a gidari.
b. A gidari aztertu epaile.

1 2 3 4



(48) a. Epaile aztertu a antzezle.
b. A antzezle aztertu epaile.

1 2 3 4

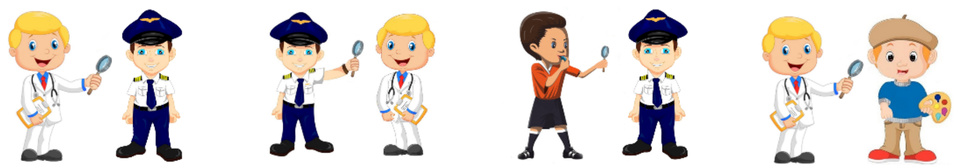


(49) a. Sendagile aztertu a margolari.
b. A margolari aztertu sendagile.

1 2 3 4

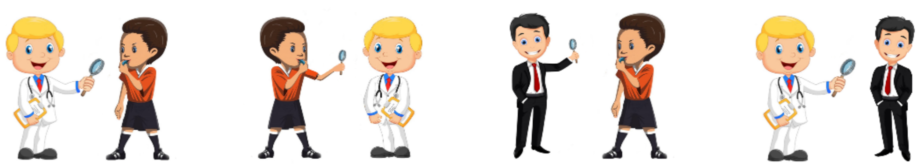
(50) a. Sendagile aztertu a gidari.
b. A gidari aztertu sendagile.

1 2 3 4



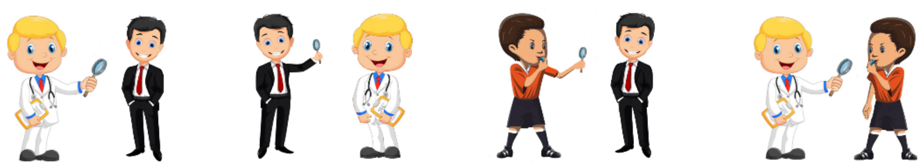
(51) a. Sendagile aztertu a epaile.
b. A epaile aztertu sendagile.

1 2 3 4



(52) a. Sendagile aztertu a antzezle.
b. A antzezle aztertu sendagile.

1 2 3 4



(53) a. Gidari aztertu a margolari.
b. A margolari aztertu gidari.

1 2 3 4

(54) a. Gidari aztertu a sendagile.
b. A sendagile aztertu gidari.

1 2 3 4

(55) a. Gidari aztertu a epaile.
b. A epaile aztertu gidari.

1 2 3 4

(56) a. Gidari aztertu a antzezle.
b. A antzezle aztertu gidari.

1 2 3 4

(57) a. Margolari aztertu a sendagile.
b. A sendagile aztertu margolari.

1 2 3 4

(58) a. Margolari aztertu a gidari.
b. A gidari aztertu margolari.

1 2 3 4

(59) a. Margolari aztertu a epaile.
b. A epaile aztertu margolari.

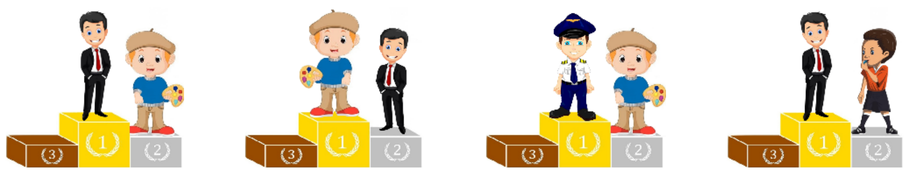
1 2 3 4

(60) a. Margolari aztertu a antzezle.
b. A antzezle aztertu margolari.

1 2 3 4

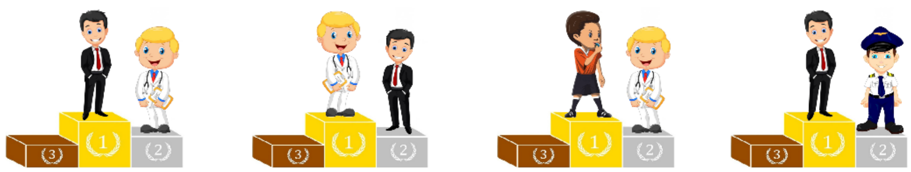
(61) a. Antzezle gainditu a margolari.
b. A margolari gainditu antzezle.

1 2 3 4



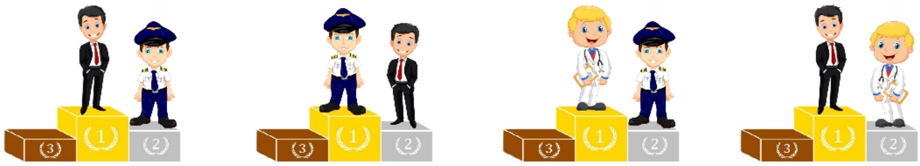
(62) a. Antzezle gainditu a sendagile.
b. A sendagile gainditu antzezle.

1 2 3 4



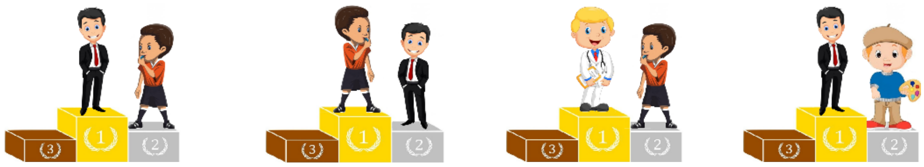
(63) a. Antzezle gainditu a gidari.
b. A gidari gainditu antzezle.

1 2 3 4



(64) a. Antzezle gainditu a epaile.
b. A epaile gainditu antzezle.

1 2 3 4

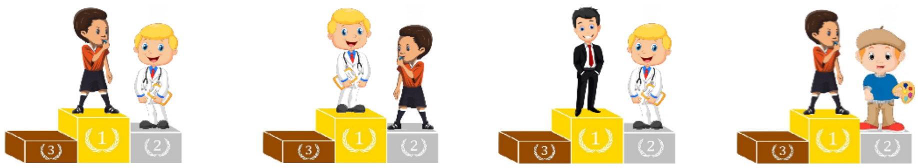


(65) a. Epaile gainditu a margolari.
b. A margolari gainditu epaile.

1 2 3 4

(66) a. Epaile gainditu a sendagile.
b. A sendagile gainditu epaile.

1 2 3 4



(67) a. Epaile gainditu a gidari.
b. A gidari gainditu epaile.

1 2 3 4

(68) a. Epaile gainditu a antzezle.
b. A antzezle gainditu epaile.

1

2

3

4


(69) a. Sendagile gainditu a margolari.
b. A margolari gainditu sendagile.

1

2

3

4


(70) a. Sendagile gainditu a gidari.
b. A gidari gainditu sendagile.

1

2

3

4


(71) a. Sendagile gainditu a epaile.
b. A epaile gainditu sendagile.

1

2

3

4


(72) a. Sendagile gainditu a antzezle.
b. A antzezle gainditu sendagile.

1

2

3

4


(73) a. Gidari gainditu a margolari.
b. A margolari gainditu gidari.

1 2 3 4

(74) a. Gidari gainditu a sendagile.
b. A sendagile gainditu gidari.

1 2 3 4

(75) a. Gidari gainditu a epaile.
b. A epaile gainditu gidari.

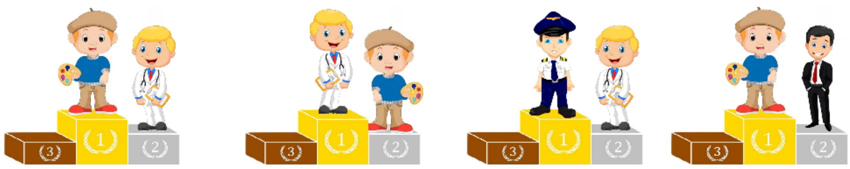
1 2 3 4

(76) a. Gidari gainditu a antzezle.
b. A antzezle gainditu gidari.

1 2 3 4

(77) a. Margolari gainditu a sendagile.
b. A sendagile gainditu margolari.

1 2 3 4



(78) a. Margolari gainditu a gidari.
b. A gidari gainditu margolari.

1

2

3

4


(79) a. Margolari gainditu a epaile.
b. A epaile gainditu margolari.

1

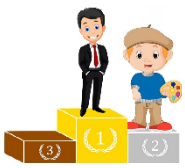
2

3

4

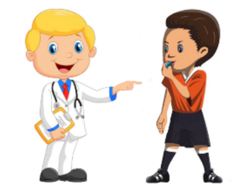

(80) a. Margolari gainditu a antzezle.
b. A antzezle gainditu margolari.

1

2

3

4


<u>Testing set: Pictures used in the production task</u>





3
1
2

9. Rubric to evaluate awareness and transcription of verbal reports in Experiment 5

<u>Rubric evaluating awareness</u>

The questions asked in the verbal report were:

1. In all the sentences you have heard there were two nouns, a verb and another word. Could you say what this other word was?
2. Could you say why it appeared in the sentence or how it was used?
3. In which part of the experiment did you realize this?
4. Did you follow any strategy to perform the test?

I present below the rubric used to evaluate awareness of patient marking based on answers to questions 1 and 2.

Participant status	**Description**
Aware	The participant states or exemplifies that the word *a* is placed in front of the patient/object of the sentence. Alternatively, s/he states or exemplifies that the word *a* is not placed in front of the agent/subject of the sentence.
Unaware	The participant is not able to identify the patient mark. Alternatively, s/he is able to identify it but cannot (correctly) say what conceptual role it marks or, more generally, what is its function.

TABLE C-9.1. Rubric used to classify participants as *aware* or *unaware* of L2 patient marking based on their responses in the verbal report of Experiment 5.

<u>Verbal reports and awareness</u>

Below I summarise cognate and non-cognate learners' responses to the questions in the verbal report. The column "Report of patient marking" corresponds to the answers to questions 1 and 2. The column "Awareness" indicates whether participants were *aware* or *unaware* of patient marking. Regarding question 3, all aware learners reported that they became aware of the function of the patient mark during the exposure phase. Finally, the column "Strategy test" reports the strategy that each participant used in the sentence-picture congruency task. Based on responses to question 4, I have classified the strategies into three categories: (i) accept syntactic conditions and reject semantic conditions (i.e. accept syntactically congruent and incongruent sentence-picture pairs and reject semantically incongruent ones), (ii) apply patient marking knowledge or (iii) intuition.

Cognate learners

Participant	Report of patient marking	Awareness	Strategy test
P1	La *a* es como una preposición. Cuando va al principio de la oración, imagino que es lo mismo que al final, pero al principio.	Unaware	Intuition
P2	La *a* es igual que la *a* en castellano. Puede ir al principio de la frase o más al final.	Aware[2]	Apply patient marking knowledge
P3	La *a* se usa para referirse a la persona a la que se dirige la acción.	Aware	Apply patient marking knowledge
P4	La *a* aparecía en dos tipos de frases, una del tipo: *El médico saluda **a**l actor* y la otra del tipo: ***A**l actor lo saluda el médico.*	Aware	Accept syntactic conditions and reject semantic conditions
P5	La *a* se usa al principio de la frase para cambiar de voz activa a voz pasiva.	Aware[3]	Apply patient marking knowledge
P6	Si la *a* está al principio o después del verbo, el sujeto que va después sufre la acción.	Aware	Apply patient marking knowledge
P7	La *a* se usa como en castellano, para decir por ejemplo: *El árbitro señala **a**l actor* o ***A**l médico lo ha espiado el actor*.	Aware	Apply patient marking knowledge
P8	Es como si dijéramos: ***A**l médico lo señala el actor* o *El árbitro señala **a**l actor*.	Aware	Apply patient marking knowledge
P9	La *a* indica hacia quién iba la acción, marca la dirección del verbo.	Aware	Accept syntactic conditions and reject semantic conditions
P10	La *a* introduce el complemento indirecto, quien recibe la acción.	Aware	Apply patient marking knowledge

[2] The participant reports that the word *a* can appear in two different positions in the sentence (i.e. in two different structures) and claims that the function of this word is the same as in Spanish, which I interpret as referring to it being a patient mark.

[3] The participant reports that when the word *a* appears at the beginning of the sentence, the sentence is in the passive voice. I interpret this as referring to the fact that the patient is in the first position of the sentence and, thus, follows *a,* and that the agent appears later in the sentence.

Participant	Report of patient marking	Awareness	Strategy test
P11	Cuando la *a* va al principio, el sujeto que va detrás es el que está afectado por la acción. Cuando la *a* está más al final, aparece primero en la oración la persona que hace la acción a la persona afectada.	Aware	Apply patient marking knowledge
P12	La *a* va delante de la persona que recibía la acción. Puede aparecer al principio de la oración o detrás del verbo.	Aware	Apply patient marking knowledge
P13	La *a* va delante de la persona a la que se le aplica el verbo o la persona que recibe la acción.	Aware	Accept syntactic conditions and reject semantic conditions
P14	Creo que la *a* va delante de quien está haciendo la acción.	Unaware	Intuition
P15	La *a* va delante de la persona a quien le hacen la acción.	Aware	Apply patient marking knowledge
P16	La *a* indica el complemento indirecto. Por ejemplo: *El pintor gana **a** X* o ***A**l pintor le están ganando*.	Aware	Intuition
P17	Creo que cuando la *a* está al principio de la oración, el que hace la acción es el segundo oficio. Cuando la *a* no va al principio, la *a* describe el sujeto que recibe la acción.	Aware	Apply patient marking knowledge
P18	La *a* se usa para decir que X persona hace algo ***a*** otra persona o que ***a*** X persona le hace algo otra persona.	Aware	Apply patient marking knowledge
P19	No tengo muy claro si la *a* es una marca de sujeto o de complemento indirecto.	Unaware	Intuition
P20	En una estructura, la primera persona hace el verbo y la que va después de la *a* es la que lo recibe. Cuando la *a* va al principio, el significado es diferente, por ejemplo: *El pintor es escogido por el árbitro*.	Aware[4]	Apply patient marking knowledge
P21	La *a* se usa como complemento directo; cuando la acción va dirigida a una persona, se pone antes.	Aware	Apply patient marking knowledge
P22	No sé por qué se usa la *a*.	Unaware	Intuition

[4] See footnote number 3.

Participant	Report of patient marking	Awareness	Strategy test
P23	Si la *a* está entre el primer y el segundo nombre, uno le hace la acción al otro. Si la *a* está al principio, el segundo le hace la acción al primero.	Aware	Apply patient marking knowledge
P24	Creo que la *a* es como un pronombre, indica quién recibía o hacía la acción.	Unaware	Accept syntactic conditions and reject semantic conditions
P25	La *a* aparece en dos tipos de frases: *El piloto ha examinado **a**l actor* o ***A**l actor lo ha examinado el piloto* / *El actor ha sido examinado por el piloto*.	Aware	Apply patient marking knowledge
P26	La *a* puede ir al principio o más hacia el final, para indicar, por ejemplo: *El actor elige **a**l pintor* o ***A**l actor lo examina el árbitro*.	Aware	Apply patient marking knowledge
P27	La *a* a veces va al principio y a veces no. En cualquier caso, la palabra que va detrás de la *a* es la persona sobre quien recaía la acción.	Aware	Apply patient marking knowledge
P28	La *a* equivale al *al* del castellano. *El actor elige **al** pintor* o ***Al** pintor lo elige el actor.*	Aware	Apply patient marking knowledge
P29	En un tipo de frases la *a* va primero, luego la persona que recibe la acción y luego la que la hace. En el otro tipo de frases, la primera persona le hace algo ***a*** la segunda.	Aware	Apply patient marking knowledge
P30	El que tiene la *a* delante es el que recibe la acción.	Aware	Apply patient marking knowledge

TABLE C-9.2. Transcription of cognate learners' responses in the verbal report of Experiment 5 (in Spanish). Based on these responses, participants are classified as *aware* or *unaware* of L2 patient marking. Each participant's reported strategy in the sentence-picture congruency task is also indicated.

English translation of verbal reports

P1: The *a* is like a preposition. I imagine that its function is the same regardless of whether it goes at the beginning of the sentence or at the end.

P2: The *a* is the same as the *a* in Spanish. It can go at the beginning of the sentence or towards the end.

P3: The *a* is used to refer to the person towards whom the action is directed.

P4: The *a* appeared in two types of sentences, one of the type: *The doctor greets* ***a*** *the actor* [SV-*a*-O order] and the other of the type: ***A*** *the actor greets the doctor* [*a*-OVS order].

P5: The *a* is used at the beginning of the sentence to change it from active voice to passive voice.

P6: If the *a* is at the beginning [of the sentence] or after the verb, the person that follows it undergoes the action.

P7: The *a* is used as in Spanish, to say, for example: *The referee points* ***a*** *the actor* [SV-*a*-O order] or ***A*** *the doctor spied on the actor* [*a*-OVS order].

P8: It is as if we were saying: ***A*** *the doctor points the actor* [*a*-OVS order] or *The referee points* ***a*** *the actor* [SV-*a*-O order].

P9: The *a* indicates who the action was directed towards; it marks the direction of the verb.

P10: The *a* introduces the indirect object, who receives the action.

P11: When the *a* is at the beginning [of the sentence], the person following it is the one affected by the action. When the *a* is towards the end, the person who does the action appears first in the sentence.

P12: The *a* goes in front of the person who receives the action. It can appear at the beginning of the sentence or after the verb.

P13: The *a* goes in front of the person to whom the verb is applied or the person who receives the action.

P14: I think the *a* goes in front of whoever is doing the action.

P15: The *a* goes in front of the person to whom the action is done.

P16: The *a* indicates the indirect object. For example: *The painter wins* ***a*** *X* [SV-*a*-O order] or ***A*** *the painter is being won* [*a*-OV order].

P17: I think that when the *a* is at the beginning of the sentence, the noun doing the action is the second one. When the *a* does not appear at the beginning, it precedes the person receiving the action.

P18: The *a* is used to say that X person does something **to** another person or that **to** another person, X person does something.

P19: I do not really know if the *a* is a subject mark or an indirect object mark.

P20: In a structure, the first person performs the verb and the person after the *a* is the one who receives it. When the *a* goes at the beginning [of the structure], the meaning is different, for example: *The painter is chosen by the referee.*

P21: The *a* is used as a direct object; it is written before the person towards whom the action is directed.

P22: I do not know why the *a* is used.

P23: If the *a* is between the first and the second noun, one does the action to the other. If the *a* is at the beginning [of the sentence], the second does the action to the first.

P24: I think the *a* is like a pronoun, it indicates who received or did the action.

P25: The *a* appears in two types of sentences: *The pilot has examined* ***a*** *the actor* [SV-*a*-O order] or ***A*** *the actor has examined the pilot* [*a*-OVS order] */ The actor has been examined by the pilot*.
P26: The *a* can go at the beginning [of the sentence] or towards the end to indicate, for example: *The actor chooses* ***a*** *the painter* [SV-*a*-O order] or ***A*** *the actor examines the referee* [*a*-OVS order].
P27: The *a* sometimes goes at the beginning [of the sentence] and sometimes it does not. In any case, the word after the *a* is the person who receives the action.
P28: The *a* is equivalent to *al* in Spanish. *The actor chooses* ***a*** *the painter* [SV-*a*-O order] or ***A*** *the painter chooses the actor* [*a*-OVS order].
P29: In one type of sentences, the *a* goes first, then the person who receives the action and then the person who does it. In the other type of sentences, the first person does something to the second.
P30: The noun preceded by *a* is the one who receives the action.

Non-cognate learners

Participant	**Report of patient marking**	**Awareness**	**Strategy test**
P1	La *a* iba siempre delante de la persona que no hacía la acción.	Aware	Apply patient marking knowledge
P2	Hay dos personas y una ejerce la acción sobre la otra. Cuando la *a* está al principio, el que ejerce la acción está a la derecha de la oración. Por ejemplo, es como si dijéramos: ***A****l médico lo examina el actor* o *El médico es señalado por el actor*. Si la *a* no está al principio, el significado es al revés, uno señala ***a***l otro.	Aware	Apply patient marking knowledge
P3	Si la *a* aparecía delante del último nombre, indicaba que el primer personaje le hacía la acción al segundo. Si aparecía al principio, indicaba que el primer personaje estaba recibiendo la acción.	Aware	Apply patient marking knowledge
P4	La *a* se usa para referirse a quién se le está llevando a cabo la acción.	Aware	Apply patient marking knowledge
P5	Solo me he fijado en que en una oración la *a* se usaba para indicar que uno hacía la acción ***a***l otro. He visto que a veces aparecía al principio, pero no sé por qué.	Aware[5]	Intuition

[5] This participant reports knowing that *a* is a patient mark, but only in SVO sentences. Thus, I consider him/her (partly) aware of patient marking.

Participant	Report of patient marking	Awareness	Strategy test
P6	Haciendo un paralelismo con el castellano, la *a* indica si el sujeto es activo o pasivo. Es decir, la *a* va delante del sujeto al que se le va a hacer la acción.	Aware	Apply patient marking knowledge
P7	La *a* me ha parecido que era como un artículo.	Unaware	Intuition
P8	La *a* se usa para generar la conexión entre el verbo y el nombre.	Unaware	Intuition
P9	La *a* marca si la persona que hace la acción está a la derecha o a la izquierda del verbo, pero no sé exactamente cómo se usa.	Unaware	Accept syntactic conditions and reject semantic conditions
P10	Cuando la frase empezaba con *a*, esa persona recibía la acción. Si estaba en el medio, la frase decía que una persona le hace la acción ***a*** otra.	Aware	Apply patient marking knowledge
P11	La *a* es una preposición. A veces va delante del verbo y otras veces, detrás. En algunas frases se usa para decir, por ejemplo, que *El árbitro examina **a**l actor* y en otras para decir que ***A**l actor le examina el piloto*.	Aware	Intuition
P12	La *a* indica la persona a la que se le hace la acción.	Aware	Apply patient marking knowledge
P13	Cuando va la *a* primero, a esa persona se le está haciendo la acción. Cuando la *a* es la tercera palabra, el sujeto va primero.	Aware	Apply patient marking knowledge
P14	Me he dado cuenta de que la *a* a veces iba delante del verbo y a veces iba detrás, pero no sé qué significaba.	Unaware	Intuition
P15	Cuando la *a* está al inicio, la oración es como una pasiva, el segundo sujeto es el que hace la acción. Cuando la *a* no va delante, el primer sujeto es el que hace la acción.	Aware	Apply patient marking knowledge
P16	La *a* indica la persona que hace o recibe la acción, pero no me ha dado tiempo de fijarme bien cuál.	Unaware	Accept syntactic conditions and reject semantic conditions

Participant	Report of patient marking	Awareness	Strategy test
P17	La palabra es la *a*. Me he dado cuenta de que había dos órdenes de palabras pero no sé qué función tiene la *a*.	Unaware	Intuition
P18	La *a* va delante de quien recibe la acción.	Aware	Apply patient marking knowledge
P19	La *a* es una preposición y se usa igual que en castellano, para decir que ***A*** X le ha hecho tal cosa Y o que X ha hecho algo ***a*** Y.	Aware	Apply patient marking knowledge
P20	Cuando la *a* va delante, la oración significa, por ejemplo: *Un actor es escogido por un piloto*. Cuando no va delante, significa, por ejemplo: *El actor elige* ***a*** *un pintor*.	Aware[6]	Intuition
P21	Cuando la *a* va al principio, indica a quién va dirigida la acción. Cuando va en medio de la frase, la primera persona es quien hace la acción.	Aware	Apply patient marking knowledge
P22	La *a* indica a quién iba dirigida la acción.	Aware	Apply patient marking knowledge.
P23	No me he dado cuenta de que hubiese ninguna palabra más en la oración.	Unaware	Intuition
P24	La *a* muestra a quién le está pasando el verbo.	Aware	Apply patient marking knowledge
P25	Dependiendo de dónde estaba la *a* era un personaje el que hacía la acción o el otro, pero no tengo muy claro cómo se usa.	Unaware	Intuition
P26	La *a* aparecía en dos estructuras. Una era, por ejemplo: *El actor elige* ***a****l pintor,* y la otra: ***A****l árbitro le elige el médico*.	Aware	Apply patient marking knowledge
P27	El verbo va dirigido hacia la persona que sigue a la *a*. Es decir, la *a* indica sobre quién recae la acción.	Aware	Apply patient marking knowledge
P28	La *a* se usa para indicar o bien sobre quién recae la acción o bien quién la hace, pero no lo tengo muy claro.	Unaware	Intuition

[6] See footnote number 3.

Participant	Report of patient marking	Awareness	Strategy test
P29	La *a* va antes de una persona cuando se refiere a la persona que recibe la acción.	Aware	Apply patient marking knowledge
P30	Solo me he dado cuenta de que la *a* se usa para decir, por ejemplo, que *El actor elige* ***a****l pintor*.	Aware[7]	Intuition

TABLE C-9.3. Transcription of non-cognate learners' responses in the verbal report of Experiment 5 (in Spanish). Based on these responses, participants are classified as *aware* or *unaware* of L2 patient marking. Each participant's reported strategy in the sentence-picture congruency task is also indicated.

English translation of verbal reports

P1: The *a* always went in front of the person who did not do the action.

P2: There are two people and one does the action to the other. When the *a* is at the beginning [of the sentence], the one who performs the action is on the right of the sentence. For example, it is as if we said: ***A*** *the doctor examines the actor* [*a*-OVS order] or *The doctor is pointed at by the actor*. If the *a* is not at the beginning [of the sentence], the meaning is the other way around.

P3: If the *a* appeared in front of the last noun, it indicated that the first character was doing the action to the second. If it appeared at the beginning [of the sentence], it indicated that the first character was receiving the action.

P4: The *a* is used to refer to the person who is receiving the action.

P5: I only noticed that in a sentence the *a* was used to indicate that a noun was doing the action **to** another noun. I have seen it sometimes appear at the beginning [of the sentence], but I do not know why.

P6: Making a parallel with Spanish, the *a* indicates whether the subject is active or passive. That is, the *a* goes in front of the person to whom the action is performed.

P7: The *a* seemed like an article.

P8: The *a* is used to generate a connection between the verb and the noun.

P9: The *a* marks whether the person doing the action is to the right or left of the verb, but I do not know exactly how it is used.

P10: When the sentence began with *a*, the person that followed received the action. If *a* was in the middle [of the sentence], the sentence meant that a person did the action **to** another.

P11: The *a* is a preposition. Sometimes it goes before the verb and sometimes, after it. In some sentences, it is used to say, for example, that *The referee examines* ***a*** *the actor* [SV-*a*-O order] and in others to say that ***A*** *the actor examines the pilot* [*a*-OVS order].

P12: The *a* indicates the person to whom the action is done.

P13: When the *a* goes first, the action is being done to the person that follows it. When the *a* is the third word, the subject goes first.

P14: I noticed that the *a* sometimes was before the verb and sometimes after it, but I do not know why.

[7] See footnote number 5.

P15: When the *a* is at the beginning [of the sentence], the sentence is like a passive, the second person is the one that does the action. When the *a* is not at the beginning of the sentence, the first person is the one that does the action.

P16: The *a* indicates the person who does or receives the action, but I have not had time to figure out which one.

P17: The word is *a*. I realized that there were two word orders but I do not know what is the function of *a*.

P18: The *a* goes before the person receiving the action.

P19: The *a* is a preposition and it is used as in Spanish, to say that: ***A*** (to) X, Y did something or X did something ***a*** (to) Y.

P20: When the *a* goes at the beginning of the sentence, this means, for example: *An actor is chosen by a pilot.* When it does not go at the beginning of the sentence, this means, for example: *The actor chooses a painter.*

P21: When the *a* goes at the beginning [of the sentence], it indicates who the action is directed to. When it goes in the middle of the sentence, the first person is the one who does the action.

P22: The *a* indicates to whom the action was directed.

P23: I did not realize that there were any other words in the sentence.

P24: The *a* shows who is receiving the verb.

P25: Depending on where the *a* was, a character did the action or the other did, but I am not very sure how it is used.

P26: The *a* appeared in two structures. One was, for example: *The actor chooses **a** the painter* [SV-*a*-O order], and the other: ***A** the painter chooses the doctor* [*a*-OVS order].

P27: The verb is directed towards the person who follows the *a*.

P28: The *a* is used to indicate either who receives the action or who does it, but I am not very sure about it.

P29: The *a* goes before the person receiving the action.

P30: I have only realized that the *a* is used to say, for example, that *The actor chooses **a** the painter* [SV-*a*-O order].

Appendix D

Resumen en español

1. Introducción

El aprendizaje de lenguas es muy relevante hoy en día, en una era de globalización en la que las personas de todo el mundo están interconectadas. En este contexto, conocer más de una lengua es muy recomendable y, en las últimas décadas, el interés por el aprendizaje de lenguas ha ido en aumento (Pauwels, 2014; Wang, 2023). En lingüística, el aprendizaje de una lengua después de que la primera lengua (L1) haya sido adquirida se conoce como *adquisición de una segunda lengua (L2)*. La adquisición de una L2, al igual que la adquisición de la L1, requiere aprender las palabras de la lengua (incluyendo su significado y pronunciación) y su sintaxis (la forma en la que las palabras se combinan para construir oraciones), entre otros aspectos. Los infantes difieren en la velocidad de adquisición, pero excepto en casos extremos en los que son privados de exposición a la lengua (e.g. Curtiss, 1977), todos alcanzan una competencia total en la L1. En cambio, la adquisición de L2 es más difícil; los aprendices varían en su ratio de adquisición y en su nivel de competencia, y tan solo unos pocos alcanzan una competencia similar a la de los hablantes nativos (R. Ellis, 2004; Housen & Simoens, 2016). En esta tesis, he investigado cómo facilitar la adquisición de L2 por parte de aprendices adultos. Me he centrado en los estadios iniciales de la adquisición, desde que los aprendices son expuestos por primera vez a la L2. De entre los muchos aspectos del lenguaje que necesitan ser adquiridos, me he centrado en la sintaxis y, específicamente, en las estructuras sintácticas, las cuales capturan las formas en las que las palabras se combinan para formar constituyentes dentro de la oración (Van Valin, 2001). Por lo tanto, en esta tesis he estudiado cómo facilitar *la adquisición inicial de la sintaxis de una L2*. Si bien este objeto de estudio se puede abordar desde muchas perspectivas, en esta tesis me he centrado en explorar la facilitación causada por dos factores lingüísticos: (i) la similitud interlingüística de la sintaxis y (ii) el procesamiento léxico.

Por una parte, hay evidencia de que la influencia interlingüística facilita el procesamiento y uso de estructuras y rasgos morfosintácticos similares en la L1 y la L2 por parte de aprendices con una competencia baja, intermedia o avanzada en la L2 (e.g. Chang & Zheng, 2015; Díaz et al., 2016; Foucart & Frenck-Mestre, 2011; Izquierdo & Collins, 2008; Tokowicz & MacWhinney, 2005). Sin embargo, una cuestión mucho menos estudiada es si la similitud interlingüística facilita, además, la adquisición de sintaxis de L2 por parte de aprendices noveles (pero véase

Tolentino & Tokowicz, 2014). Por otra parte, algunas palabras se procesan más fácilmente que otras. Por ejemplo, las *palabras de alta frecuencia,* i.e. palabras que se usan a menudo en la lengua, se procesan más rápido y con mayor acierto que las *palabras de baja frecuencia,* i.e. palabras que se usan menos en la lengua (Cop et al., 2015; Duyck et al., 2008; Gollan et al., 2008, 2011; Whitford & Titone, 2012 y más). Esta ventaja de procesamiento también ocurre para los *cognados,* i.e. palabras con forma y significado parecido en dos o más lenguas (e.g. *tomato* [inglés] – *tomaat* [neerlandés]) en comparación a los *no cognados,* i.e. palabras con significado parecido pero forma diferente en dos o más lenguas (e.g. *tomato* [inglés] – *pomodoro* [italiano]) (Costa et al., 2000; Dijkstra et al., 1999, 2010; Lemhöfer & Dijkstra, 2004; Van Assche et al., 2011 y más). Estudios previos sugieren que las palabras de alta frecuencia (Hopp, 2016; Tily et al., 2010) y los cognados (X. Chen et al., 2023; Hopp, 2017) facilitan el procesamiento sintáctico en la L1 y la L2. Sin embargo, no se ha investigado si el procesamiento de palabras de alta vs. baja frecuencia y cognados vs. no cognados afecta la adquisición de estructuras sintácticas de la L2 y, de ser así, cómo.

Postular hipótesis sobre cómo la similitud interlingüística de la sintaxis y el procesamiento léxico descrito podrían afectar la adquisición inicial de estructuras de la L2 requiere un amplio entendimiento de la mente bilingüe. Más concretamente, es necesario comprender cómo el lexicón y la sintaxis de la L1 y la L2 (incluyendo palabras que varían en frecuencia y en su estatus cognado-no cognado, estructuras interlingüísticamente similares y disimilares) se representan e interactúan durante la adquisición de L2 en tiempo real. A mi entender, no existe un modelo, teoría o marco teórico que especifique explícitamente todos estos aspectos (pero varias teorías tratan algunos aspectos por separado, ver Capítulo 1). Sin embargo, hay un marco teórico que proporciona una visión detallada acerca de cómo las lenguas se representan, procesan y adquieren en tiempo real y que es compatible con múltiples de las teorías mencionadas: el MOGUL (por sus siglas en inglés, *Modular On-line Growth and Use of Language*, Sharwood Smith, 2017; Sharwood Smith & Truscott, 2014). Este es el marco teórico en el cual me he basado para proponer las hipótesis de la tesis doctoral.

En este trabajo, presento cinco experimentos conductuales donde estudio cómo aprendices noveles de L2 adquieren estructuras similares vs. disimilares a la L1 (ver Capítulo 2) y cómo la adquisición de este tipo de estructuras se ve afectada por la frecuencia léxica (ver Capítulo 3) y el estatus cognado o no cognado de las palabras (ver Capítulo 4). En la siguiente sección, resumo el contenido de los capítulos experimentales.

2. Resumen de los capítulos experimentales

2.1. Capítulo 2. El rol facilitador de la similitud sintáctica interlingüística en la adquisición inicial de sintaxis de la L2

En el Capítulo 2, he investigado si la similitud entre una estructura de la L2 y su equivalente en la L1 facilita la adquisición por parte de aprendices noveles. Varias teorías y modelos de adquisición de L2 predicen esta facilitación. Estos asumen, implícita o explícitamente, que las estructuras que son interlingüísticamente similares se procesan usando representaciones sintácticas de la L1 que forman parte del sistema lingüístico del aprendiz desde el inicio del proceso de adquisición de la L2. Por el contrario, las estructuras que son interlingüísticamente disimilares deben ser adquiridas en base al input (Carroll, 1999, 2001; MacWhinney, 2005; B. D. Schwartz & Sprouse, 1994, 1996; Westergaard, 2021). El Experimento 1 tenía como objetivo comparar cómo hablantes nativos de español sin conocimiento de gallego aprendían dos estructuras subordinadas de esta lengua, una similar y la otra disimilar a las estructuras subordinadas equivalentes del español. El experimento empezaba con una fase de exposición que consistía en una tarea auditiva de juicios de plausibilidad. Los participantes eran expuestos a oraciones plausibles e implausibles con vocabulario cognado, la mitad formadas por la estructura interlingüísticamente similar y la otra mitad formadas por la estructura interlingüísticamente disimilar. Después, una tarea auditiva de juicios de gramaticalidad testeaba el aprendizaje de las estructuras, reflejado en la habilidad de distinguir la estructura similar y la disimilar de sus agramaticales. Los resultados de esta tarea sugirieron que los aprendices conocían la estructura similar, pero no mostraron evidencia de que la estructura disimilar se hubiese aprendido.

En el Experimento 2, hice algunos cambios en el diseño experimental, transformando el paradigma de aprendizaje de implícito a explícito. Más concretamente, (i) cambié la tarea de juicios de plausibilidad por una tarea de búsqueda de estructuras en el input, (ii) doblé el número de oraciones formadas por la estructura similar y la disimilar en la fase de exposición, (iii) presenté las oraciones tanto auditivamente como por escrito en dicha exposición y en el test y (iv) incluí *feedback* en la tarea de juicios de gramaticalidad (un tick verde o una cruz roja indicaban si las respuestas de los participantes eran correctas o incorrectas, respectivamente). En este caso, el test reveló que tanto la estructura similar como la disimilar formaban parte del sistema lingüístico de los aprendices y que el aprendizaje era significativamente mayor para la estructura similar que para la estructura disimilar. Este resultado fue interpretado como evidencia a favor de la facilitación de la similitud sintáctica interlingüística en el aprendizaje de estructuras de la L2. Por otra parte, en el Experimento 2 también abordé dos preguntas de investigación metodológicas. En primer lugar, me pregunté si la estructura disimilar se había aprendido durante la fase de exposición o durante el test, gracias al *feeedback*. Para responder a esta pregunta, se analizaron los primeros 20 ítems del test, que reflejaban el aprendizaje inmediatamente después de la exposición. En estos primeros ítems, los aprendices ya eran capaces de distinguir la estructura disimilar de su agramatical, lo cual sugirió que la estructura disimilar se había aprendido durante la exposición. En segundo lugar, me pregunté acerca del efecto del *feedback* en la adquisición de las estructuras. Estudios previos han mostrado que el *feedback* puede facilitar el aprendizaje de la sintaxis de la L2 (e.g. Leeman, 2003; Mackey & Philp, 1998; Muranoi, 2000; Rosa, 1999). De la misma manera, en el Experimento 2 el aprendizaje de las estructuras observado en los primeros 20 ítems del test aumentó al llegar a los últimos 20 ítems. En los primeros ítems, el aprendizaje de la estructura similar ya era mayor que el de la estructura disimilar, y esto no cambió al final del test.

Por último, tanto en el Experimento 1 como en el Experimento 2 pregunté a los participantes si podían verbalizar las estructuras para así medir el conocimiento consciente o inconsciente de las mismas que resultó del aprendizaje implícito (Experimento 1) o explícito (Experimento 2). Puesto que estas preguntas se hicieron al final de todos los experimentos de la tesis, resumiré de forma general las conclusiones extraídas una vez presentados todos los experimentos.

2.2. Capítulo 3. El rol facilitador de la frecuencia léxica en la adquisición inicial de sintaxis de la L2

En el Capítulo 3, se ha investigado si el hecho de procesar palabras que varían en activación debido a su frecuencia de uso afecta la adquisición de estructuras de la L2 que tienen o no tienen una estructura equivalente en la L1. Tal y como se ha mencionado, la influencia de la frecuencia léxica en la adquisición de sintaxis de la L2 no se ha explorado y tan solo un par de estudios han mostrado que esta puede influir en el procesamiento de estructuras de la L1 y la L2 (Hopp, 2016; Tily et al., 2010). Para que las diferencias de frecuencia léxica puedan afectar la adquisición de sintaxis de la L2, los aprendices deben haber sido expuestos a la L2 y deben haber procesado palabras con diferente frecuencia, de tal manera que algunas tengan una mayor frecuencia que otras. Sin embargo, en esta tesis he estudiado el aprendizaje de L2 por parte de adultos que nunca habían sido expuestos a la segunda lengua. En el Capítulo 3, he propuesto que la frecuencia léxica puede afectar la adquisición de estructuras sintácticas de la L2 siempre y cuando las palabras que contienen la manipulación de frecuencia sean cognadas en la L1 y la L2 y la frecuencia manipulada sea la de las palabras en la L1. Teniendo esto en cuenta, realicé el Experimento 3, el cual replicaba el Experimento 2 pero usando verbos cognados entre el español y el gallego con una frecuencia en español (la L1 de los participantes) significativamente más baja que la de los verbos del Experimento 2. Asumí que, tanto en el Experimento 2 como en el Experimento 3, cuando los participantes procesaran los verbos cognados sus equivalentes en español se activarían y los verbos del gallego se procesarían como si fueran simplemente las traducciones de los verbos del español. De esta forma, la mayor o menor frecuencia de los verbos del español se asumiría también para los verbos del gallego. Consecuentemente, la comparación de los resultados de los Experimentos 2 y 3 permitiría determinar el rol de la frecuencia léxica en la adquisición de estructuras de la L2 similares o disimilares a la L1. Se postularon tres hipótesis al respecto.

En primer lugar, se postuló la hipótesis de que la mayor activación de los verbos de alta frecuencia (Experimento 2) en comparación a los verbos de baja frecuencia (Experimento 3) facilitaría la adquisición de la estructura interlingüísticamente disimilar. Más concretamente, basándome en el MOGUL propuse que, cuanto mayor fuera la activación de una palabra, mayor sería el nivel de activación de la estructura que incluía esa palabra durante el procesamiento y mayor sería el nivel de activación en reposo de la estructura cuando el procesamiento terminara. En términos del MOGUL, esto significa que su aprendizaje también sería mayor. La comparación de los resultados de los tests (tarea de juicios de gramaticalidad) de los Experimentos 2 y 3 confirmó esta hipótesis, puesto que mostró que el aprendizaje de la estructura disimilar era significativamente mayor cuando se aprendía con verbos de alta frecuencia que con verbos de baja frecuencia. En segundo lugar, se postuló la hipótesis de que el efecto facilitador de la frecuencia léxica en la adquisición de sintaxis no se obtendría para la estructura de la L2 similar a la L1, la cual se procesaría usando una estructura de la L1 ya aprendida y, por lo tanto, se vería menos afectada por diferencias en la frecuencia léxica. En términos del MOGUL, inicialmente esta estructura ya tendría un nivel de activación en reposo alto y, por lo tanto, este variaría menos a causa del procesamiento de la estructura con verbos de alta o baja frecuencia. Esta hipótesis también se confirmó, puesto que el aprendizaje de la estructura similar no difirió entre el Experimento 2 y el Experimento 3. Por último, la tercera hipótesis dictaba que el mayor aprendizaje de la estructura similar en comparación a la disimilar observado en el Experimento 2 se replicaría en el Experimento 3. No obstante, se esperaba que la diferencia entre el aprendizaje de las dos estructuras fuera mayor cuando estas se aprendían con verbos de baja frecuencia que con verbos de alta frecuencia. Como evidencia a favor de esta hipótesis, el aprendizaje de la estructura similar fue mayor que el de la estructura disimilar en ambos experimentos, pero la magnitud del efecto fue mayor en el Experimento 3 que en el Experimento 2.

2.3. Capítulo 4. El rol facilitador de los cognados en la adquisición inicial de sintaxis de la L2

En el Capítulo 4, he investigado si el hecho de procesar palabras que varían en activación debido a su estatus cognado o no cognado afecta la adquisición de estructuras de la L2 disimilares (Experimento 4) o similares (Experimento 5) a la L1. Tal y como he comentado con anterioridad, la influencia de los cognados en la adquisición de sintaxis de la L2 no ha sido investigada y solo unos pocos estudios han mostrado que estas palabras pueden facilitar el procesamiento sintáctico (e.g. X. Chen et al., 2023; Hopp, 2017; J. Huang et al., 2019; Soares et al., 2018, 2019). Creé dos versiones de una mini-lengua basada en euskera: una con nombres no cognados entre el euskera y el español y verbos cognados en las dos lenguas (*versión cognada*) y la otra con nombres y verbos del euskera no cognados con el español (*versión no cognada*). En el Experimento 4, las estructuras a aprender tenían un orden SOV u OSV y marcaban el agente y el paciente de la oración mediante posposiciones. Estas estructuras, basadas en la gramática del euskera, no existen en español. En el Experimento 5, las estructuras tenían un orden SVO u OVS y marcaban el paciente de la oración mediante una preposición. Esta vez, las estructuras eran iguales al español.

En cada experimento, dos grupos de hablantes nativos de español sin conocimiento de euskera aprendieron el vocabulario de la versión cognada o la versión no cognada de la lengua mediante dibujos, juntamente con sus traducciones en la L1. El aprendizaje del vocabulario se testeó mediante dos tareas. Una tarea consistía en relacionar las palabras con sus dibujos y la otra en nombrar dichos dibujos. Los participantes repitieron las tareas tantas veces como fue necesario hasta alcanzar el 100% de aciertos. Tal y como se muestra en estudios previos

(Antón & Duñabeitia, 2020; Comesaña et al., 2019; Marecka et al., 2021; Valente et al., 2018, y más), los Experimentos 4 y 5 indicaron que los cognados se aprenden más fácilmente que los no cognados. Los participantes que aprendieron cognados alcanzaron el 100% de aciertos para estos verbos en una o ambas tareas en menos intentos que los participantes que aprendieron no cognados. Varios modelos de aprendizaje léxico pueden explicar este efecto (e.g. Grainger et al., 2010; Hall, 2002; Kroll & Stewart, 1994). Este resultado se debe a la similitud interlingüística entre los cognados y sus traducciones en la L1, la cual podría haber causado que los cognados de la L2 se establecieran en el sistema lingüístico de los aprendices más fácilmente y más firmemente que los no cognados.

Tras aprender el vocabulario, los dos grupos fueron expuestos a oraciones formadas por las dos estructuras de la L2 de forma implícita, esto es, sin saber que se trataba de dos tipos de oraciones diferentes. Los verbos de las oraciones eran cognados para un grupo de participantes, mientras que para el otro eran no cognados. Todas las oraciones iban acompañadas por un dibujo que representaba su significado. Los participantes debían observar cada dibujo y escuchar y leer la oración que lo describía. Tras la fase de exposición, todos los participantes aprendieron nuevos verbos no cognados, los cuales se usaron más tarde en dos tareas que testeaban el aprendizaje de las estructuras. Una tarea consistía en detectar la congruencia o incongruencia entre pares de oraciones y dibujos (es decir, detectar si la oración describía correctamente el dibujo al que acompañaba o no). Para hacer esto, los participantes debían prestar atención a las marcas de agente y/o paciente en la oración y a los nombres del agente y del paciente en la misma. La otra tarea consistía en escribir oraciones para describir dibujos de acciones transitivas usando las estructuras y el vocabulario aprendido. Por un lado, la primera tarea no mostró resultados fiables, puesto que en los dos experimentos algunos aprendices juzgaron la congruencia o incongruencia entre los pares de oraciones y dibujos usando únicamente su conocimiento léxico, y no el sintáctico. Por otro lado, en la segunda tarea aquellos participantes que aprendieron las estructuras de la L2 con verbos cognados mostraron un uso de las estructuras más acertado que aquellos participantes que aprendieron las estructuras con verbos no cognados, pero solo cuando estas eran interlingüísticamente disimilares (Experimento 4), no sucedió lo mismo cuando eran interlingüísticamente similares (Experimento 5). Esto sugiere que los cognados facilitaron el aprendizaje de las estructuras de la L2 disimilares a la L1, pero no facilitaron el aprendizaje de las estructuras de la L2 similares a la L1. Como en el Capítulo 3, estos resultados fueron interpretados en base al marco teórico MOGUL. Argumenté que la mayor activación de los cognados comparado con los no cognados se expandió a las estructuras de la L2, produciendo un nivel de activación mayor de las mismas durante el procesamiento y un nivel de activación en reposo mayor después del procesamiento. Esto se tradujo en un mayor aprendizaje de las estructuras procesadas con verbos cognados que de las estructuras procesadas con verbos no cognados, pero solamente cuando estas estructuras no podían ser procesadas usando una estructura equivalente de la L1 ya aprendida y con un nivel de activación en reposo elevado.

Finalmente, voy a resumir los resultados de las preguntas que midieron el conocimiento consciente o inconsciente de la sintaxis de la L2 derivado de los experimentos. El aprendizaje implícito de sintaxis de la L2 suele resultar en conocimiento inconsciente o no verbalizable (e.g. Kim & Fenn, 2020; Leung & Williams, 2006; Rebuschat, 2009; Tagarelli et al., 2016; Williams, 2005). Por el contrario, el aprendizaje explícito de sintaxis de la L2 suele producir conocimiento consciente o verbalizable (e.g. N. C. Ellis, 1993; Rebuschat, 2009; Robinson, 1997; Tagarelli et al., 2016). Mis estudios de aprendizaje explícito (Experimentos 2 y 3) coincidieron con los estudios previos, puesto que dieron como resultado conocimiento verbalizable de las estructuras de la L2 para la mayoría de los participantes. En cambio, dejando de lado el Experimento 1, el cual no mostró un aprendizaje claro de las estructuras, mis estudios de aprendizaje implícito (Experimentos 4 y 5) también ocasionaron conocimiento verbalizable de la sintaxis de la L2 para un gran número de participantes. Argumenté que este resultado podía ser consecuencia de la tarea de exposición, en la cual, si bien no se indicó a los aprendices que tenían que centrarse en la forma de las oraciones, tampoco se desvió la atención de dicha forma. Además, los aprendices adultos tienen consciencia metalingüística, la cual les podría haber conducido a buscar regularidades en el input.

3. Conclusiones generales

En esta tesis he investigado la facilitación ejercida por la similitud sintáctica interlingüística y el procesamiento léxico en la adquisición inicial de estructuras de la L2, tanto similares como disimilares a la L1. Las contribuciones principales de la tesis son:

1. He mostrado que los aprendices adultos noveles de L2 obtienen un mayor aprendizaje de las estructuras que también existen en la L1 que de aquellas estructuras que solo existen en la L2 (Capítulos 2 y 3). Esto constituye nueva evidencia a favor del rol facilitador de la similitud sintáctica interlingüística en el estadio más temprano del desarrollo de la L2 y valida los modelos y teorías de adquisición de L2 que predicen esta facilitación.

2. He mostrado por primera vez que la adquisición de estructuras de la L2 disimilares a la L1 es mejor cuando estas incluyen palabras que obtienen una mayor activación durante el procesamiento: palabras de alta frecuencia, en comparación a las de baja frecuencia (Capítulo 3), y cognados, en comparación a los no cognados (Capítulo 4). Por otro lado, he mostrado que la adquisición de estructuras de la L2 similares a la L1 es igual (ni mejor ni peor) cuando estas incluyen palabras de alta frecuencia y de baja frecuencia (Capítulo 3), cognados y no cognados (Capítulo 4). Estos resultados respaldan la hipótesis de que el procesamiento léxico, y en particular la activación léxica, facilita la adquisición de estructuras de la L2, pero que esta facilitación se ve modulada por la similitud sintáctica interlingüística.

3. He extendido las aportaciones del MOGUL respecto a la representación, procesamiento y adquisición del lexicón y la sintaxis de la L1 y la L2 en la mente bilingüe. Más concretamente, he propuesto hipótesis acerca de cómo el MOGUL podría dar cuenta de la representación, procesamiento y/o adquisición de estructuras interlingüísticamente similares y disimilares, palabras que varían en frecuencia y estatus cognado o no cognado, y la influencia que el procesamiento de estas palabras tendría en la adquisición de estructuras de la L2 similares y disimilares a la L1 (Capítulos 1 al 4, ver Capítulo 5 para un resumen). Esto constituye un esfuerzo teórico considerable, puesto que ninguna teoría o modelo de adquisición de L2 trata explícitamente todos estos aspectos.

4. Futuras líneas de investigación

En esta tesis he investigado dos aspectos de la adquisición de sintaxis de L2 por parte de adultos. Por un lado, he explorado la influencia positiva de la similitud sintáctica interlingüística en la adquisición inicial de estructuras sintácticas, la cual había sido poco estudiada. Los estudios presentados en los Capítulos 2 y 3 amplían el entendimiento de esta cuestión. Por otro lado, el efecto facilitador de las palabras de alta frecuencia y de los cognados en la adquisición inicial de estructuras interlingüísticamente similares o disimilares no se había investigado. Los experimentos en los Capítulos 3 y 4 constituyen un primer paso hacia la comprensión de este tema. Los resultados de esta tesis deben ser replicados en futuras investigaciones, puesto que reproducir los resultados de la investigación experimental es de vital importancia para fortalecer su validez (Open Science Collaboration, 2015). A continuación, se discuten tres futuras líneas de investigación adicionales.

4.1. Evaluar las implicaciones pedagógicas de los resultados de la tesis

En esta tesis he investigado cómo se puede facilitar la adquisición inicial de sintaxis de la L2 por parte de aprendices adultos, un aspecto que es altamente relevante en una era de globalización donde aprender lenguas es muy importante. Aunque mi trabajo pertenece al campo de la psicolingüística, considero que mis resultados pueden ser de interés para campos como la enseñanza de segundas lenguas y pueden promover la investigación interdisciplinaria. Se podrían intentar extraer implicaciones pedagógicas de los resultados reportados y sugerir, por ejemplo, que los profesores usen cognados, palabras de alta frecuencia y estructuras interlingüísticamente similares para facilitar el aprendizaje sintáctico en el aula. Sin embargo, hay que tener en cuenta que los resultados presentados en esta tesis han sido obtenidos mediante experimentos que se han llevado a cabo en un laboratorio y bajo condiciones altamente controladas, por lo cual no es evidente que se pudieran obtener los mismos resultados en un contexto de aprendizaje diferente. Las técnicas y herramientas de enseñanza usadas en el aula difieren de las usadas en mis experimentos psicolingüísticos. Para evaluar las implicaciones pedagógicas reales de los resultados de esta tesis, sería necesario diseñar experimentos que recreasen las condiciones de aprendizaje en el aula, para así testear si el rol facilitador de la similitud sintáctica interlingüística y el procesamiento léxico observado también se obtiene en esas circunstancias.

4.2. Explorar el efecto facilitador de la similitud sintáctica interlingüística y el procesamiento léxico más allá de los estadios iniciales de la adquisición de sintaxis de la L2

En esta tesis he estudiado la adquisición de estructuras sintácticas en el estadio más inicial del desarrollo de la L2. Todos los experimentos se han realizado con adultos que no habían sido expuestos nunca a las lenguas en las cuales se basaban las L2s estudiadas y su primer encuentro con estas lenguas fue en el laboratorio. Puesto que el foco de mi estudio eran los estadios iniciales del aprendizaje, realicé experimentos cortos, llevados a cabo en una sesión y testeando el aprendizaje tras una exposición relativamente breve a las estructuras de la L2 (en todos los experimentos, la fase de exposición duraba 10 minutos o menos y la duración total de los experimentos oscilaba entre los 45 minutos y la hora y media). Futuras investigaciones podrían explorar si la facilitación ejercida por la similitud sintáctica interlingüística, las palabras de alta frecuencia y los cognados en la adquisición de estructuras de la L2 varía a medida que los aprendices reciben más exposición a las estructuras, tal y como ocurriría en el curso natural del aprendizaje de una L2. Para ello, estudios futuros podrían replicar los experimentos de esta tesis, pero añadiendo una segunda sesión experimental realizada en un segundo día. En esta segunda sesión, los participantes serían expuestos de nuevo a las estructuras de la L2 y, seguidamente, serían testeados acerca de su conocimiento sintáctico. Mi hipótesis es que la facilitación que resulta de la similitud sintáctica interlingüística y el procesamiento léxico desaparecería gradualmente en cuanto los aprendices fueran alcanzando una mayor competencia en la L2.

4.3. Seguir investigando el efecto facilitador de la activación léxica en la adquisición de estructuras interlingüísticamente disimilares

Una de las conclusiones principales de esta tesis es que la activación léxica puede facilitar la adquisición de estructuras disimilares en la L1 y la L2. Futuros estudios deberían obtener más evidencia empírica a favor de esta afirmación. Por ejemplo, se podría replicar el Experimento 4 pero usando una mini-lengua que incluyera cognados ortográficamente idénticos en dos o más lenguas (e.g. *pintar* [español] – *pintar* [catalán]) en vez de cognados no idénticos, con ortografía y/o fonología ligeramente diferente en dos lenguas (e.g. *pintar* [español] – *pintatu* [euskera]). Los bilingües procesan más rápido aquellos cognados con un mayor grado de similitud ortográfica (y/o fonológica) que aquellos con un menor grado de similitud (e.g. Dijkstra et al., 2010; Duyck et al., 2007; Van Assche et al., 2011). Esto se explica por el hecho de que a medida que la similitud del cognado en dos lenguas aumenta, el grado de activación interlingüística también incrementa, y esto causa una mayor facilitación en el procesamiento. Por lo tanto, la mayor co-activación y, consecuentemente, la mayor facilitación, se obtiene para los cognados que tienen exactamente la misma forma en dos lenguas (Dijkstra et al., 2010). Si, tal y como he propuesto en esta tesis, la mayor activación de los cognados no idénticos en comparación a los no cognados facilitó el aprendizaje de estructuras interlingüísticamente disimilares en el Experimento 4, entonces esta facilitación podría aumentar si el estudio se replicase con cognados idénticos.

Por último, otra opción sería replicar los Experimentos 2 y 3, realizados con verbos de alta y baja frecuencia, respectivamente, pero usando una manipulación de frecuencia más extrema. Si bien la frecuencia media era significativamente mayor para los verbos de alta frecuencia que para los verbos de baja frecuencia, en algunos casos individuales la diferencia entre un verbo de alta frecuencia y uno de baja frecuencia no era muy pronunciada. Los estudios que investigan los efectos de frecuencia léxica en el procesamiento a menudo usan palabras de alta y/o baja frecuencia con un alto grado de variabilidad (e.g. en Lehtonen et al., 2012, rango de alta frecuencia de 7.89–504 observaciones por millón y rango de baja frecuencia de 0.04–4.23 observaciones por millón, ver también Gollan et al., 2008). Sin embargo, si tal y como he propuesto en el Capítulo 3, la mayor activación de los verbos de alta frecuencia en comparación a los de baja frecuencia facilitó la adquisición de estructuras interlingüísticamente disimilares, entonces es posible que la facilitación aumente si la diferencia en frecuencia y, por lo tanto, en activación entre los dos grupos de verbos es aún más grande. Para testear esta hipótesis, futuros estudios podrían replicar los Experimentos 2 y 3 pero estableciendo rangos de alta y baja frecuencia de entre los cuales se seleccionarían verbos con menor variabilidad que en los Experimentos 2 y 3.

References

Abrahamsson, N., & Hyltenstam, K. (2008). The robustness of aptitude effects in near-native second language acquisition. *Studies in Second Language Acquisition*, *30*(4), 481–509. https://doi.org/10.1017/S027226310808073X

Alamer, A., & Alrabai, F. (2023). The causal relationship between learner motivation and language achievement: New dynamic perspective. *Applied Linguistics*, *44*(1), 148–168. https://doi.org/10.1093/applin/amac035

Alamer, A., & Lee, J. (2019). A motivational process model explaining L2 Saudi students' achievement of English. *System*, *87*, 102133. https://doi.org/10.1016/j.system.2019.102133

Alanen, R. (1995). Input enhancement and rule presentation in second language acquisition. In R. Schmidt (Ed.), *Attention and awareness in foreign language learning* (pp. 259–302). University of Hawai'i Press.

Aldezabal, I., Aranzabe, M., Atutxa, A., Gojenola, K., Sarasola, K., & Zabala, I. (2003). *Hitz-hurrenkeraren azterketa masiboa corpusean*.

Alonso, R. (2016). *Crosslinguistic influence in second language acquisition*. Multilingual Matters.

Alrabai, F. (2022). The predictive role of anxiety and motivation in L2 proficiency: An empirical causal model. *Language Teaching Research*, 1–39. https://doi.org/10.1177/13621688221136247

Altmann, G., & Steedman, M. (1988). Interaction with context during human sentence processing. *Cognition*, *30*(3), 191–238. https://doi.org/10.1016/0010-0277(88)90020-0

Andras, F., Rivera, M., Bajo, T., Dussias, P. E., & Paolieri, D. (2022). Cognate facilitation effect during auditory comprehension of a second language: A visual world eye-tracking study. *International Journal of Bilingualism*, *26*(4), 405–425. https://doi.org/10.1177/13670069211033359

Antón, E., & Duñabeitia, J. A. (2020). Better to be alone than in bad company: Cognate synonyms impair word learning. *Behavioral Sciences*, *10*(8), 123. https://doi.org/10.3390/bs10080123

Arai, M., van Gompel, R. P. G., & Scheepers, C. (2007). Priming ditransitive structures in comprehension. *Cognitive Psychology*, *54*(3), 218–250. https://doi.org/10.1016/j.cogpsych.2006.07.001

Arıbaş, D. Ş., & Cele, F. (2021). Acquisition of articles in L2 and L3 English: The influence of L2 proficiency on positive transfer from L2 to L3. *Journal of Multilingual and Multicultural Development*, *42*(1), 19–36. https://doi.org/10.1080/01434632.2019.1667364

Baayen, R. H., Piepenbrock, R., & Gulikers, R. (1995). *Celex2*. Linguistic Data Consortium. https://doi.org/10.35111/GS6S-GM48

Balling, L. W. (2013). Reading authentic texts: What counts as cognate? *Bilingualism: Language and Cognition*, *16*(3), 637–653. https://doi.org/10.1017/S1366728911000733

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology:*

Human Perception and Performance, *10*(3), 340–357. https://doi.org/10.1037/0096-1523.10.3.340

Balota, D. A., Pilotti, M., & Cortese, M. J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Memory & Cognition*, *29*(4), 639–647. https://doi.org/10.3758/BF03200465

Barber, H. A., Otten, L. J., Kousta, S.-T., & Vigliocco, G. (2013). Concreteness in word processing: ERP and behavioral effects in a lexical decision task. *Brain and Language*, *125*(1), 47–53. https://doi.org/10.1016/j.bandl.2013.01.005

Bardovi-Harlig, K. (1997). Another piece of the puzzle: The emergence of the present perfect. *Language Learning*, *47*(3), 375–422. https://doi.org/10.1111/0023-8333.00015

Bardovi-Harlig, K., & Sprouse, R. A. (2017). Negative versus positive transfer. In J. I. Liontas & M. DelliCarpini (Eds.), *The TESOL encyclopedia of English language teaching* (pp. 1–6). John Wiley & Sons. https://doi.org/10.1002/9781118784235.eelt0084

Bastarrika, A., & Davidson, D. J. (2017). An event related field study of rapid grammatical plasticity in adult second-language learners. *Frontiers in Human Neuroscience*, *11*, 12. https://doi.org/10.3389/fnhum.2017.00012

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). *Parsimonious mixed models*. arXiv.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Bates, E., & MacWhinney, B. (1981). Second-language acquisition from a functionalist perspective: Pragmatic, semantic, and perceptual strategies. *Annals of the New York Academy of Sciences*, *379*(1), 190–214. https://doi.org/10.1111/j.1749-6632.1981.tb42009.x

Bates, E., & MacWhinney, B. (1982). Functionalist approaches to grammar. In E. Wanner & L. Gleitman (Eds.), *Language acquisition: The state of the art* (pp. 173–218). Cambridge University Press.

Bernolet, S., Hartsuiker, R. J., & Pickering, M. J. (2007). Shared syntactic representations in bilinguals: Evidence for the role of word-order repetition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(5), 931–949. https://doi.org/10.1037/0278-7393.33.5.931

Bernolet, S., Hartsuiker, R. J., & Pickering, M. J. (2009). Persistence of emphasis in language production: A cross-linguistic approach. *Cognition*, *112*(2), 300–317. https://doi.org/10.1016/j.cognition.2009.05.013

Bernolet, S., Hartsuiker, R. J., & Pickering, M. J. (2012). Effects of phonological feedback on the selection of syntax: Evidence from between-language syntactic priming. *Bilingualism: Language and Cognition*, *15*(3), 503–516. https://doi.org/10.1017/S1366728911000162

Berwick, R. C., & Chomsky, N. (2016). *Why only us: Language and evolution*. The MIT Press.

Besner, D., & McCann, R. S. (1987). Word frequency and pattern distortion in visual word identification and production. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 201–209). Lawrence Erlbaum Associates.

Bialystok, E., & Hakuta, K. (1999). Confounded age: Linguistic and cognitive factors in age differences for second language acquisition. In D. Birdsong (Ed.), *Second language acquisition and the Critical period hypothesis* (pp. 161–181). Lawrence Erlbaum Associates.

Bijeljac-babic, R., Biardeau, A., & Grainger, J. (1997). Masked orthographic priming in bilingual word recognition. *Memory & Cognition*, *25*(4), 447–457. https://doi.org/10.3758/BF03201121

Birdsong, D. (1992). Ultimate attainment in second language acquisition. *Language*, *68*(4), 706–755. https://doi.org/10.1353/lan.1992.0035

Birdsong, D., & Molis, M. (2001). On the evidence for maturational constraints in second-language acquisition. *Journal of Memory and Language*, *44*(2), 235–249. https://doi.org/10.1006/jmla.2000.2750

Blosfeld, M. E., & Bradley, D. C. (1981). *Visual and auditory word recognition: Effects of frequency and syllabicity [Paper presentation]*. Third Australian Language and Speech Conference, Melbourne, Australia.

Blumenfeld, H. K., & Marian, V. (2007). Constraints on parallel activation in bilingual spoken language processing: Examining proficiency and lexical status using eye-tracking. *Language and Cognitive Processes*, *22*(5), 633–660. https://doi.org/10.1080/01690960601000746

Blumenfeld, H. K., & Marian, V. (2013). Parallel language activation and cognitive control during spoken word recognition in bilinguals. *Journal of Cognitive Psychology*, *25*(5), 547–567. https://doi.org/10.1080/20445911.2013.812093

Boersma, P., & Weenink, D. (2018). *Praat: Doing phonetics by computer* (6.0.37).

Bokander, L. (2020). Language aptitude and crosslinguistic influence in initial L2 learning. *Journal of the European Second Language Association*, *4*(1), 35–44. https://doi.org/10.22599/jesla.69

Bolhuis, J. J., Okanoya, K., & Scharff, C. (2010). Twitter evolution: Converging mechanisms in birdsong and human speech. *Nature Reviews Neuroscience*, *11*(11), 747–759. https://doi.org/10.1038/nrn2931

Bosque, I., & Gutiérrez-Rexach, J. (2009). *Fundamentos de sintaxis formal*. Ediciones AKAL.

Brysbaert, M., Lagrou, E., & Stevens, M. (2017). Visual word recognition in a second language: A test of the lexical entrenchment hypothesis with lexical decision times. *Bilingualism: Language and Cognition*, *20*(3), 530–548. https://doi.org/10.1017/S1366728916000353

Brysbaert, M., Mandera, P., & Keuleers, E. (2018). The word frequency effect in word processing: An updated review. *Current Directions in Psychological Science*, *27*(1), 45–50. https://doi.org/10.1177/0963721417727521

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990. https://doi.org/10.3758/BRM.41.4.977

Brysbaert, M., Stevens, M., Mandera, P., & Keuleers, E. (2016). The impact of word prevalence on lexical decision times: Evidence from the Dutch Lexicon Project 2. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(3), 441–458. https://doi.org/10.1037/xhp0000159

Bultena, S., Dijkstra, T., & van Hell, J. G. (2014). Cognate effects in sentence context depend on word class, L2 proficiency, and task. *Quarterly Journal of Experimental Psychology*, *67*(6), 1214–1241. https://doi.org/10.1080/17470218.2013.853090

Bybee, J. (2008). Usage-based grammar and second language acquisition. In P. Robinson & N. C. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 216–237). Routledge.

Cai, Z. G., Pickering, M. J., Yan, H., & Branigan, H. P. (2011). Lexical and syntactic representations in closely related languages: Evidence from Cantonese–Mandarin bilinguals. *Journal of Memory and Language*, *65*(4), 431–445. https://doi.org/10.1016/j.jml.2011.05.003

Carreiras, M., Gernsbacher, M. A., & Villa, V. (1995). The advantage of first mention in Spanish. *Psychonomic Bulletin & Review*, *2*(1), 124–129. https://doi.org/10.3758/BF03214418

Carreiras, M., Perea, M., & Grainger, J. (1997). Effects of the orthographic neighborhood in visual word recognition: Cross-task comparisons. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*(4), 857–871. https://doi.org/10.1037/0278-7393.23.4.857

Carroll, S. (1999). Putting 'input' in its proper place. *Second Language Research*, *15*(4), 337–388. https://doi.org/10.1191/026765899674928444

Carroll, S. (2001). *Input and evidence: The raw material of second language acquisition*. John Benjamins Publishing Company.

Carroll, S., & Swain, M. (1993). Explicit and implicit negative feedback: An empirical study of the learning of linguistic generalizations. *Studies in Second Language Acquisition*, *15*(3), 357–386. https://doi.org/10.1017/S0272263100012158

Casaponsa, A., Antón, E., Pérez, A., & Duñabeitia, J. A. (2015). Foreign language comprehension achievement: Insights from the cognate facilitation effect. *Frontiers in Psychology*, *06*, 588. https://doi.org/10.3389/fpsyg.2015.00588

Casenhiser, D., & Goldberg, A. E. (2005). Fast mapping between a phrasal form and meaning. *Developmental Science*, *8*(6), 500–508. https://doi.org/10.1111/j.1467-7687.2005.00441.x

Centro Ramón Piñeiro para a investigación en humanidades. (2022). *Corpus de Referencia do Galego Actual (CORGA)*. http://corpus.cirp.gal/corga/

Chai, X., & Bao, J. (2023). Linguistic distances between native languages and Chinese influence acquisition of Chinese character, vocabulary, and grammar. *Frontiers in Psychology*, *13*, 1083574. https://doi.org/10.3389/fpsyg.2022.1083574

Chang, H., & Zheng, L. (2015). The role of L1 in the acquisition of Chinese causative constructions by English-speaking learners. In L. Yu & T. Odlin (Eds.), *New perspectives on transfer in second language learning* (pp. 145–168). Multilingual Matters.

Chaudron, C. (1985). Intake: On models and methods for discovering learners' processing of input. *Studies in Second Language Acquisition*, *7*(1), 1–14. https://doi.org/10.1017/S027226310000512X

Chen, B., Jia, Y., Wang, Z., Dunlap, S., & Shin, J.-A. (2013). Is word-order similarity necessary for cross-linguistic structural priming? *Second Language Research*, *29*(4), 375–389. https://doi.org/10.1177/0267658313491962

Chen, X., Wang, S., & Hartsuiker, R. J. (2023). Do structure predictions persevere to multilinguals' other languages? Evidence from cross-linguistic structural priming in comprehension. *Bilingualism: Language and Cognition*, *26*(4), 653–669. https://doi.org/10.1017/S1366728922000888

Cherici, A. (2023). The role of L1 and L2 in the acquisition of null subjects by Chinese learners of L3 Italian. *International Journal of Multilingualism*, *20*(3), 735–752. https://doi.org/10.1080/14790718.2021.1977308

Chomsky, N. (1965). *Aspects of the theory of syntax*. The MIT Press. https://www.jstor.org/stable/j.ctt17kk81z

Chomsky, N. (1972). *Language and mind* (Enlarged ed). Harcourt Brace Jovanovich.

Chomsky, N. (1980). *Rules and representations*. Columbia University Press.

Chomsky, N. (1982). *Some concepts and consequences of the theory of government and binding*. MIT Press.

Chomsky, N. (1995). *The Minimalist program*. The MIT Press.

Cintrón-Valentín, M. C., & Ellis, N. C. (2016). Salience in second language acquisition: Physical form, learner attention, and instructional focus. *Frontiers in Psychology*, *7*, 1284. https://doi.org/10.3389/fpsyg.2016.01284

Clifton, C., & Frazier, L. (1989). Comprehending sentences with long-distance dependencies. In G. N. Carlson & M. K. Tanenhaus (Eds.), *Linguistic structure in language processing* (pp. 273–317). Kluwer Academic Publishers.

Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. Academic Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed). Lawrence Erlbaum Associates.

Colina, S. (2009). *Spanish phonology: A syllabic perspective*. Georgetown University Press.

Comesaña, M., Bertin, P., Oliveira, H., Soares, A. P., Hernández-Cabrera, J. A., & Casalis, S. (2018). The impact of cognateness of word bases and suffixes on morpho-orthographic processing: A masked priming study with intermediate and high-proficiency Portuguese-English bilinguals. *PLOS ONE*, *13*(3), e0193480. https://doi.org/10.1371/journal.pone.0193480

Comesaña, M., Ferré, P., Romero, J., Guasch, M., Soares, A. P., & García-Chico, T. (2015). Facilitative effect of cognate words vanishes when reducing the orthographic overlap: The role of stimuli list composition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(3), 614–635. https://doi.org/10.1037/xlm0000065

Comesaña, M., Moreira, A. J., Valente, D., Hernández-Cabrera, J. A., & Soares, A. P. (2019). List composition effect on cognate and non-cognate word acquisition in children. *Linguistic Approaches to Bilingualism*, *9*(2), 289–313. https://doi.org/10.1075/lab.16034.com

Comesaña, M., Sánchez-Casas, R., Soares, A. P., Pinheiro, A. P., Rauber, A., Frade, S., & Fraga, I. (2012). The interplay of phonology and orthography in visual cognate word recognition: An ERP study. *Neuroscience Letters*, *529*(1), 75–79. https://doi.org/10.1016/j.neulet.2012.09.010

Comesaña, M., Soares, A. P., Sánchez-Casas, R., & Lima, C. (2012). Lexical and semantic representations in the acquisition of L2 cognate and non-cognate words: Evidence from two learning methods in children. *British Journal of Psychology*, *103*(3), 378–392. https://doi.org/10.1111/j.2044-8295.2011.02080.x

Connine, C. M., Mullennix, J., Shernoff, E., & Yelen, J. (1990). Word familiarity and frequency in visual and auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(6), 1084–1096. https://doi.org/10.1037/0278-7393.16.6.1084

Cop, U., Dirix, N., Van Assche, E., Drieghe, D., & Duyck, W. (2017). Reading a book in one or two languages? An eye movement study of cognate facilitation in L1 and L2 reading. *Bilingualism: Language and Cognition*, *20*(4), 747–769. https://doi.org/10.1017/S1366728916000213

Cop, U., Keuleers, E., Drieghe, D., & Duyck, W. (2015). Frequency effects in monolingual and bilingual natural reading. *Psychonomic Bulletin & Review*, *22*(5), 1216–1234. https://doi.org/10.3758/s13423-015-0819-2

Corder, S. P. (1967). The significance of learner's errors. *International Review of Applied Linguistics in Language Teaching*, *5*(1–4), 161–170. https://doi.org/10.1515/iral.1967.5.1-4.161

Costa, A., Caramazza, A., & Sebastián-Gallés, N. (2000). The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1283–1296. https://doi.org/10.1037/0278-7393.26.5.1283

Cuetos, F., González-Nosti, M., Barbón, A., & Brysbaert, M. (2011). SUBTLEX-ESP: Spanish word frequencies based on film subtitles. *Psicológica*, *32*(2), 133–143.

Cuetos, F., & Mitchell, D. C. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the Late Closure strategy in Spanish. *Cognition*, *30*(1), 73–105. https://doi.org/10.1016/0010-0277(88)90004-2

Curtiss, S. (1977). *Genie: A psycholinguistic study of a modern-day wild child*. Elsevier. https://doi.org/10.1016/C2013-0-07305-7

Davies, M. (2015). *Corpus of contemporary American English (COCA)*. Harvard Dataverse. https://doi.org/10.7910/DVN/AMUDUW

de Graaff, R. (1997). The eXperanto experiment: Effects of explicit instruction on second language acquisition. *Studies in Second Language Acquisition*, *19*(2), 249–276. https://doi.org/10.1017/S0272263197002064

de Groot, A. M. B., Borgwaldt, S., Bos, M., & Van Den Eijnden, E. (2002). Lexical decision and word naming in bilinguals: Language effects and task effects. *Journal of Memory and Language*, *47*(1), 91–124. https://doi.org/10.1006/jmla.2001.2840

de Groot, A. M. B., & Comijs, H. (1995). Translation recognition and translation production: Comparing a new and an old tool in the study of bilingualism. *Language Learning*, *45*(3), 467–509. https://doi.org/10.1111/j.1467-1770.1995.tb00449.x

de Groot, A. M. B., & Keijzer, R. (2000). What is hard to learn is easy to forget: The roles of word concreteness, cognate status, and word frequency in foreign-language vocabulary learning and forgetting. *Language Learning*, *50*(1), 1–56. https://doi.org/10.1111/0023-8333.00110

de Groot, A. M. B., & Nas, G. L. J. (1991). Lexical representation of cognates and noncognates in compound bilinguals. *Journal of Memory and Language*, *30*(1), 90–123. https://doi.org/10.1016/0749-596X(91)90012-9

De Rijk, R. (1969). Is Basque an SOV language? *Fontes Linguae Vasconum: Studia et Documenta*, *1*(3), 319–351.

De Rijk, R. (2007). *Standard Basque: A progressive grammar*. The MIT Press. https://doi.org/10.7551/mitpress/7444.001.0001

DeKeyser, R. M. (1995). Learning second language grammar rules: An experiment with a miniature linguistic system. *Studies in Second Language Acquisition*, *17*(3), 379–410. https://doi.org/10.1017/S027226310001425X

DeKeyser, R. M. (2003). Implicit and explicit learning. In C. J. Doughty & H. M. Long (Eds.), *The handbook of second language acquisition* (pp. 312–348). Blackwell.

Dekydtspotter, L. (2001). The Universal Parser and interlanguage: Domain-specific mental organization in the comprehension of combien interrogatives in English–French interlanguage. *Second Language Research*, *17*(2), 91–143. https://doi.org/10.1191/026765801678646233

Del Maschio, N., & Abutalebi, J. (2019). Language organization in the bilingual and multilingual brain. In J. W. Schwieter & M. Paradis (Eds.), *The handbook of the neuroscience of* *multilingualism* (pp. 197–213). John Wiley & Sons. https://doi.org/10.1002/9781119387725.ch9

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321. https://doi.org/10.1037/0033-295X.93.3.283

Dell, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes*, *5*(4), 313–349. https://doi.org/10.1080/01690969008407066

Denhovska, N., Serratrice, L., & Payne, J. (2016). Acquisition of second language grammar under incidental learning conditions: The role of frequency and working memory. *Language Learning*, *66*(1), 159–190. https://doi.org/10.1111/lang.12142

Díaz, B., Erdocia, K., De Menezes, R. F., Mueller, J. L., Sebastián-Gallés, N., & Laka, I. (2016). Electrophysiological correlates of second-language syntactic processes are related to native and second language distance regardless of age of acquisition. *Frontiers in Psychology*, *7*, 133. https://doi.org/10.3389/fpsyg.2016.00133

Dienes, Z., & Longuet-Higgins, C. (2004). Can musical transformations be implicitly learned? *Cognitive Science*, *28*(4), 531–558. https://doi.org/10.1207/s15516709cog2804_2

Dijkstra, T., Grainger, J., & van Heuven, W. J. B. (1999). Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of Memory and Language*, *41*(4), 496–518. https://doi.org/10.1006/jmla.1999.2654

Dijkstra, T., Miwa, K., Brummelhuis, B., Sappelli, M., & Baayen, H. (2010). How cross-language similarity and task demands affect cognate recognition. *Journal of Memory and Language*, *62*(3), 284–301. https://doi.org/10.1016/j.jml.2009.12.003

Dijkstra, T., & van Heuven, W. J. B. (1998). The BIA model and bilingual word recognition. In J. Grainger & A. Jacobs (Eds.), *Localist connectionist approaches to human cognition* (pp. 189–225). Lawrence Erlbaum Associates.

Dijkstra, T., & van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, *5*(3), 175–197. https://doi.org/10.1017/S1366728902003012

Dijkstra, T., Van Jaarsveld, H., & Brinke, S. T. (1998). Interlingual homograph recognition: Effects of task demands and language intermixing. *Bilingualism: Language and Cognition*, *1*(1), 51–66. https://doi.org/10.1017/S1366728998000121

Doughty, C., & Varela, E. (1998). Communicative focus on form. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 114–138). Cambridge University Press.

Drummond, A. (2013). *Ibex Farm*. http://spellout.net/ibexfarm/

Dryer, M. S. (2013). Order of subject, object and verb. In M. S. Dryer & M. Haspelmath (Eds.), *The world atlas of language structures online*. Max Planck Institute for Evolutionary Anthropology.

Dufour, S., Brunellière, A., & Frauenfelder, U. H. (2013). Tracking the time course of word-frequency effects in auditory word recognition with event-related potentials. *Cognitive Science*, *37*(3), 489–507. https://doi.org/10.1111/cogs.12015

Dupoux, E., & Mehler, J. (1990). Monitoring the lexicon with normal and compressed speech: Frequency effects and the prelexical code. *Journal of Memory and Language*, *29*(3), 316–335. https://doi.org/10.1016/0749-596X(90)90003-I

Dussias, P. E., & Cramer Scaltz, T. R. (2008). Spanish–English L2 speakers' use of subcategorization bias information in the resolution of temporary ambiguity during second language reading. *Acta Psychologica*, *128*(3), 501–513. https://doi.org/10.1016/j.actpsy.2007.09.004

Duyck, W., Van Assche, E., Drieghe, D., & Hartsuiker, R. J. (2007). Visual word recognition by bilinguals in a sentence context: Evidence for nonselective lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(4), 663–679. https://doi.org/10.1037/0278-7393.33.4.663

Duyck, W., Vanderelst, D., Desmet, T., & Hartsuiker, R. J. (2008). The frequency effect in second-language visual word recognition. *Psychonomic Bulletin & Review*, *15*(4), 850–855. https://doi.org/10.3758/PBR.15.4.850

Ecke, P., & Hall, C. J. (1998). Tres niveles de la representación mental: Evidencia de errores léxicos en estudiantes de un tercer idioma. *Estudios de Lingüística Aplicada*, *28*, 15–26.

Ellis, C., Conradie, S., & Huddlestone, K. (2013). The acquisition of grammatical gender in L2 German by learners with Afrikaans, English or Italian as their L1. *Stellenbosch Papers in Linguistics*, *41*, 17–27. https://doi.org/10.5774/41-0-131

Ellis, N. C. (1993). Rules and instances in foreign language learning: Interactions of explicit and implicit knowledge. *European Journal of Cognitive Psychology*, *5*(3), 289–318. https://doi.org/10.1080/09541449308520120

Ellis, N. C., & Beaton, A. (1993). Psycholinguistic determinants of foreign language vocabulary learning. *Language Learning*, *43*(4), 559–617. https://doi.org/10.1111/j.1467-1770.1993.tb00627.x

Ellis, R. (2004). Individual differences in second language learning. In A. Davies & C. Elder (Eds.), *The handbook of applied linguistics* (pp. 525–551). John Wiley & Sons. https://doi.org/10.1002/9780470757000.ch21

Ellis, R. (2009). Implicit and explicit learning, knowledge and instruction. In H. Reinders, R. Erlam, J. Philp, S. Loewen, & C. Elder (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 3–26). Multilingual Matters. https://doi.org/10.21832/9781847691767-003

Epstein, S., Flynn, S., & Martohardjono, G. (1996). Second language acquisition: Theoretical and experimental issues in contemporary research. *Behavioral and Brain Sciences*, *19*(4), 677–714. https://doi.org/10.1017/S0140525X00043521

Epstein, S., Flynn, S., & Martohardjono, G. (1998). The strong continuity hypothesis: Some evidence concerning functional categories in adult L2 acquisition. In S. Flynn, G. Martohardjono, & W. O'Neil (Eds.), *The generative study of second language acquisition* (pp. 61–77). Lawrence Erlbaum Associates.

Erdocia, K., Laka, I., Mestres-Missé, A., & Rodriguez-Fornells, A. (2009). Syntactic complexity and ambiguity resolution in a free word order language: Behavioral and electrophysiological evidences from Basque. *Brain and Language*, *109*(1), 1–17. https://doi.org/10.1016/j.bandl.2008.12.003

Eubank, L. (1993). On the transfer of parametric values in L2 development. *Language Acquisition*, *3*(3), 183–208. https://doi.org/10.1207/s15327817la0303_1

Eubank, L. (1994). Optionality and the initial state in L2 development. In T. Hoekstra & B. D. Schwartz (Eds.), *Language acquisition studies in generative grammar* (pp. 369–388). John Benjamins Publishing Company.

Eubank, L. (1996). Negation in early German-English interlanguage: More valueless features in the L2 initial state. *Second Language Research*, *12*(1), 73–106. https://doi.org/10.1177/026765839601200104

Euskaltzaindia. (2016). *Euskaltzaindiaren hiztegia* (2nd ed.). Euskaltzaindia.

Fábregas, A. (2013). Differential object marking in Spanish: State of the art. *Borealis – An International Journal of Hispanic Linguistics*, *2*(2), 1–80. https://doi.org/10.7557/1.2.2.2603

Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, *25*(3), 348–368. https://doi.org/10.1016/0749-596X(86)90006-9

Ferreira, F., & Henderson, J. M. (1991). How is verb information used during syntactic parsing? In G. B. Simpson (Ed.), *Understanding word and sentence* (pp. 305–330). Elsevier Science Publishers.

Flege, J. E. (1995). Second language speech learning: Theory, findings and problems. In W. Stange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). York Press.

Fleischer, Z., Pickering, M. J., & Mclean, J. F. (2012). Shared information structure: Evidence from cross-linguistic priming. *Bilingualism: Language and Cognition*, *15*(3), 568–579. https://doi.org/10.1017/S1366728911000551

Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. MIT Press.

Fodor, J. A. (2000). *The mind doesn't work that way: The scope and limits of computational psychology*. MIT Press.

Ford, M. (1983). A method for obtaining measures of local parsing complexity throughout sentences. *Journal of Verbal Learning and Verbal Behavior*, *22*(2), 203–218. https://doi.org/10.1016/S0022-5371(83)90156-1

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior*, *12*(6), 627–635. https://doi.org/10.1016/S0022-5371(73)80042-8

Forster, K. I., & Shen, D. (1996). No enemies in the neighborhood: Absence of inhibitory neighborhood effects in lexical decision and semantic categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(3), 696–713. https://doi.org/10.1037/0278-7393.22.3.696

Foucart, A., & Frenck-Mestre, C. (2011). Grammatical gender processing in L2: Electrophysiological evidence of the effect of L1–L2 syntactic similarity. *Bilingualism: Language and Cognition*, *14*(3), 379–399. https://doi.org/10.1017/S136672891000012X

Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (Third edition). SAGE.

Frazier, L. (1987). Theories of sentence processing. In J. Gardfield (Ed.), *Modularity in knowledge representation and natural-language processing* (pp. 291–307). MIT Press.

Frazier, L. (1989). Against lexical generation of syntax. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 505–528). MIT Press.

Frazier, L., & Clifton, C. (1997). Construal: Overview, motivation, and some new evidence. *Journal of Psycholinguistic Research*, *26*(3), 277–295. https://doi.org/10.1023/A:1025024524133

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, *6*(4), 291–325. https://doi.org/10.1016/0010-0277(78)90002-1

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, *14*(2), 178–210. https://doi.org/10.1016/0010-0285(82)90008-1

Fricke, M. (2022). Modulation of cross-language activation during bilingual auditory word recognition: Effects of language experience but not competing background noise. *Frontiers in Psychology*, *13*, 674157. https://doi.org/10.3389/fpsyg.2022.674157

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, *6*(2), 78–84. https://doi.org/10.1016/S1364-6613(00)01839-8

García-Miguel, J. M., Vaamonde, G., & González Domínguez, F. (2010). ADESSE. A database with syntactic and semantic annotation of a corpus of Spanish. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, M. Rosner, & D. Tapias (Eds.), *Proceedings of the seventh international conference on language resources and evaluation* (pp. 1903–1910). European Language Resources Association.

Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes*, *12*(5–6), 613–656. https://doi.org/10.1080/016909697386646

Gass, S. M. (1997). *Input, interaction, and the second language learner*. Lawrence Erlbaum Associates.

Gennari, S. P., & MacDonald, M. C. (2008). Semantic indeterminacy in object relative clauses. *Journal of Memory and Language*, *58*(2), 161–187. https://doi.org/10.1016/j.jml.2007.07.004

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*(1), 1–76. https://doi.org/10.1016/S0010-0277(98)00034-1

Gimenes, M., & New, B. (2016). Worldlex: Twitter and blog word frequencies for 66 languages. *Behavior Research Methods*, *48*(3), 963–972. https://doi.org/10.3758/s13428-015-0621-0

Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford University Press.

Goldberg, A. E., Casenhiser, D., & Sethuraman, N. (2004). Learning argument structure generalizations. *Cognitive Linguistics*, *15*(3), 289–316. https://doi.org/10.1515/cogl.2004.011

Goldberg, A. E., Casenhiser, D., & White, T. R. (2007). Constructions as categories of language. *New Ideas in Psychology*, *25*(2), 70–86. https://doi.org/10.1016/j.newideapsych.2007.02.004

Gollan, T. H., Fennema-Notestine, C., Montoya, R. I., & Jernigan, T. L. (2007). The bilingual effect on Boston Naming Test performance. *Journal of the International Neuropsychological Society*, *13*(2), 197–208. https://doi.org/10.1017/S1355617707070038

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, *58*(3), 787–814. https://doi.org/10.1016/j.jml.2007.07.001

Gollan, T. H., Slattery, T. J., Goldenberg, D., Van Assche, E., Duyck, W., & Rayner, K. (2011). Frequency drives lexical access in reading but not in speaking: The frequency-lag hypothesis. *Journal of Experimental Psychology: General*, *140*(2), 186–209. https://doi.org/10.1037/a0022256

Gómez-Vidal, B., Arantzeta, M., Laka, J. P., & Laka, I. (2022). Subjects are not all alike: Eye-tracking the agent preference in Spanish. *PLOS ONE*, *17*(8), e0272211. https://doi.org/10.1371/journal.pone.0272211

Gordon, B., & Caramazza, A. (1985). Lexical access and frequency sensitivity: Frequency saturation and open/closed class equivalence. *Cognition*, *21*(2), 95–115. https://doi.org/10.1016/0010-0277(85)90047-2

Gordon, P. C., Hendrick, R., & Johnson, M. (2001). Memory interference during language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(6), 1411–1423. https://doi.org/10.1037/0278-7393.27.6.1411

Gordon, P. C., Hendrick, R., & Levine, W. H. (2002). Memory-load interference in syntactic processing. *Psychological Science*, *13*(5), 425–430. https://doi.org/10.1111/1467-9280.00475

Grainger, J., & Dijkstra, T. (1992). On the representation and use of language information in bilinguals. *Advances in Psychology*, *83*, 207–220. https://doi.org/10.1016/S0166-4115(08)61496-X

Grainger, J., Midgley, K., & Holcomb, P. (2010). Re-thinking the bilingual interactive-activation model from a developmental perspective (BIA-d). In M. Kail & M. Hickmann (Eds.), *Language acquisition across linguistic and cognitive systems* (pp. 267–283). John Benjamins Publishing Company.

Grainger, J., Muneaux, M., Farioli, F., & Ziegler, J. C. (2005). Effects of phonological and orthographic neighbourhood density interact in visual word recognition. *The Quarterly Journal of Experimental Psychology*, *58*(6), 981–998. https://doi.org/10.1080/02724980443000386

Granena, G. (2013). Individual differences in sequence learning ability and second language acquisition in early childhood and adulthood. *Language Learning*, *63*(4), 665–703. https://doi.org/10.1111/lang.12018

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. John Wiley & Sons.

Griffin, Z. M., & Bock, K. (1998). Constraint, word frequency, and the relationship between lexical processing levels in spoken word production. *Journal of Memory and Language*, *38*(3), 313–338. https://doi.org/10.1006/jmla.1997.2547

Guo, Y. (2022). From a simple to a complex aspectual system: Feature reassembly in L2 acquisition of Chinese imperfective markers by English speakers. *Second Language Research*, *38*(1), 89–116. https://doi.org/10.1177/0267658320911433

Gut, U. (2010). Cross-linguistic influence in L3 phonological acquisition. *International Journal of Multilingualism*, *7*(1), 19–38. https://doi.org/10.1080/14790710902972248

Gutiérrez, R. (2008). La identificación de los tópicos y los focos. *Nueva Revista de Filología Hispánica (NRFH)*, *56*(2), 363–401. https://doi.org/10.24201/nrfh.v56i2.969

Hall, C. J. (1996). La estrategia parasítica: Un modelo psicolingüístico del aprendizaje de vocabulario. In S. Cuevas & J. Haidar (Eds.), *La imaginación y la intelegencia en el lenguaje: Homenaje a Roman Jakobson* (pp. 229–238). Instituto Nacional de Antropología e Historia.

Hall, C. J. (2002). The automatic cognate form assumption: Evidence for the parasitic model of vocabulary development. *IRAL - International Review of Applied Linguistics in Language Teaching*, *40*(2), 69–87. https://doi.org/10.1515/iral.2002.008

Hammarberg, B. (2001). Roles of L1 and L2 in L3 production and acquisition. In J. Cenoz, B. Hufeisen, & U. Jessner (Eds.), *Cross-linguistic influence in third language acquisition* (pp. 21–41). Multilingual Matters. https://doi.org/10.21832/9781853595509-003

Hartsuiker, R. J., Beerts, S., Loncke, M., Desmet, T., & Bernolet, S. (2016). Cross-linguistic structural priming in multilinguals: Further evidence for shared syntax. *Journal of Memory and Language*, *90*, 14–30. https://doi.org/10.1016/j.jml.2016.03.003

Hartsuiker, R. J., & Bernolet, S. (2017). The development of shared syntax in second language learning. *Bilingualism: Language and Cognition*, *20*(2), 219–234. https://doi.org/10.1017/S1366728915000164

Hartsuiker, R. J., & Pickering, M. J. (2008). Language integration in bilingual sentence production. *Acta Psychologica*, *128*(3), 479–489. https://doi.org/10.1016/j.actpsy.2007.08.005

Hartsuiker, R. J., Pickering, M. J., & Veltkamp, E. (2004). Is syntax separate or shared between languages? : Cross-linguistic syntactic priming in Spanish-English bilinguals. *Psychological Science*, *15*(6), 409–414. https://doi.org/10.1111/j.0956-7976.2004.00693.x

Hauk, O., & Pulvermüller, F. (2004). Effects of word length and frequency on the human event-related potential. *Clinical Neurophysiology*, *115*(5), 1090–1103. https://doi.org/10.1016/j.clinph.2003.12.020

Henry, N., Culman, H., & VanPatten, B. (2009). More on the effects of explicit information in instructed SLA. *Studies in Second Language Acquisition*, *31*(04), 559–575. https://doi.org/10.1017/S0272263109990027

Hogg, R. V., Tanis, E. A., & Zimmerman, D. L. (2015). *Probability and statistical inference* (Ninth edition). Pearson.

Hopp, H. (2014). Working memory effects in the L2 processing of ambiguous relative clauses. *Language Acquisition*, *21*(3), 250–278. https://doi.org/10.1080/10489223.2014.892943

Hopp, H. (2016). The timing of lexical and syntactic processes in second language sentence comprehension. *Applied Psycholinguistics*, *37*(5), 1253–1280. https://doi.org/10.1017/S0142716415000569

Hopp, H. (2017). Cross-linguistic lexical and syntactic co-activation in L2 sentence processing. *Linguistic Approaches to Bilingualism*, *7*(1), 96–130. https://doi.org/10.1075/lab.14027.hop

Hopp, H. (2018). The bilingual mental lexicon in L2 sentence processing. *Second Language*, *17*, 5–27. https://doi.org/10.11431/secondlanguage.17.0_5

Hopp, H., Steinlen, A., Schelletter, C., & Piske, T. (2019). Syntactic development in early foreign language learning: Effects of L1 transfer, input, and individual factors. *Applied Psycholinguistics*, *40*(05), 1241–1267. https://doi.org/10.1017/S0142716419000249

Hornstein, N., & Lightfoot, D. (1981). Introduction. In N. Hornstein & D. Lightfoot (Eds.), *Explanation in linguistics: The logical problem of language acquisition* (pp. 9–31). Longman.

Housen, A., & Simoens, H. (2016). Introduction: Cognitive perspectives on difficulty and complexity in L2 acquisition. *Studies in Second Language Acquisition*, *38*(2), 163–175. https://doi.org/10.1017/S0272263116000176

Howes, D. H., & Solomon, R. L. (1951). Visual duration threshold as a function of word-probability. *Journal of Experimental Psychology*, *41*(6), 401–410. https://doi.org/10.1037/h0056020

Hsieh, Y. (2017). Structural priming during sentence comprehension in Chinese–English bilinguals. *Applied Psycholinguistics*, *38*(3), 657–678. https://doi.org/10.1017/S0142716416000382

Hualde, J. I. (1992). *Catalan*. Routledge.

Hualde, J. I. (2000). Patterns of correspondence in the adaptation of Spanish borrowings in Basque. In S. Chang, L. Liaw, & J. Ruppenhofer (Eds.), *Proceedings of the twenty-fifth annual meeting of the Berkeley linguistics society* (Vol. 25, pp. 348–358). Berkeley Linguistics Society. https://doi.org/10.3765/bls.v25i1.1196

Hualde, J. I., & Ortiz de Urbina, J. (Eds.). (2003). *A grammar of Basque*. Mouton de Gruyter.

Huang, J., Pickering, M. J., Chen, X., Cai, Z., Wang, S., & Branigan, H. P. (2019). Does language similarity affect representational integration? *Cognition*, *185*, 83–90. https://doi.org/10.1016/j.cognition.2019.01.005

Huang, Y., & Ferreira, F. (2020). The application of signal detection theory to acceptability judgments. *Frontiers in Psychology*, *11*, 73. https://doi.org/10.3389/fpsyg.2020.00073

Hudson, P. T. W., & Bergman, M. W. (1985). Lexical knowledge in word recognition: Word length and word frequency in naming and lexical decision tasks. *Journal of Memory and Language*, *24*(1), 46–58. https://doi.org/10.1016/0749-596X(85)90015-4

Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics*, *40*(6), 431–439. https://doi.org/10.3758/BF03208203

Izquierdo, J., & Collins, L. (2008). The facilitative role of L1 influence in tense-aspect marking: A comparison of Hispanophone and Anglophone learners of French. *The Modern Language Journal*, *92*(3), 350–368. https://doi.org/10.1111/j.1540-4781.2008.00751.x

Jackendoff, R. (1987). *Consciousness and the computational mind*. MIT Press.

Jackendoff, R. (1997). *The architecture of the language faculty*. MIT Press.

Jackendoff, R. (1999). The representational structures of the language faculty and their interactions. In C. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 37–70). Oxford University Press.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford University Press.

Jackson, C. (2008). Proficiency level and the interaction of lexical and morphosyntactic information during L2 sentence processing. *Language Learning*, *58*(4), 875–909. https://doi.org/10.1111/j.1467-9922.2008.00481.x

Jacob, G., & Felser, C. (2016). Reanalysis and semantic persistence in native and non-native garden-path recovery. *Quarterly Journal of Experimental Psychology*, *69*(5), 907–925. https://doi.org/10.1080/17470218.2014.984231

Jacob, G., Katsika, K., Family, N., & Allen, S. E. M. (2017). The role of constituent order and level of embedding in cross-linguistic structural priming. *Bilingualism: Language and Cognition*, *20*(2), 269–282. https://doi.org/10.1017/S1366728916000717

Jarvis, S., & Pavlenko, A. (2008). *Crosslinguistic influence in language and cognition*. Routledge.

Juffs, A. (1998). Main verb versus reduced relative clause ambiguity resolution in L2 sentence processing. *Language Learning*, *48*(1), 107–147. https://doi.org/10.1111/1467-9922.00034

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*(4), 329–354. https://doi.org/10.1037/0033-295X.87.4.329

Kachinske, I., Osthus, P., Solovyeva, K. S., & Long, M. (2015). Implicit learning of a L2 morphosyntactic rule, and its relevance for language teaching. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 387–415). John Benjamins Publishing Company.

Kallini, J., Papadimitriou, I., Futrell, R., Mahowald, K., & Potts, C. (2024). *Mission: Impossible language models*. arXiv. http://arxiv.org/abs/2401.06416

Kantola, L., & van Gompel, R. P. G. (2011). Between- and within-language priming is the same: Evidence for shared bilingual syntactic representations. *Memory & Cognition*, *39*(2), 276–290. https://doi.org/10.3758/s13421-010-0016-5

Kassambara, A. (2021). *Rstatix: Pipe-friendly framework for basic statistical tests* (0.7.0). https://CRAN.R-project.org/package=rstatix

Kellerman, E., & Sharwood Smith, M. (Eds.). (1986). *Crosslinguistic influence in second language acquisition* (1st ed). Pergamon Institute of English.

Kidd, E., Tennant, E., & Nitschke, S. (2015). Shared abstract representation of linguistic structure in bilingual sentence comprehension. *Psychonomic Bulletin & Review*, *22*(4), 1062–1067. https://doi.org/10.3758/s13423-014-0775-2

Kim, K. M., & Fenn, K. M. (2020). Sleep-dependent consolidation of second language grammar knowledge: The role of awareness. *Studies in Second Language Acquisition*, *42*(5), 1107–1120. https://doi.org/10.1017/S0272263120000200

Kinsella, C., & Singleton, D. (2014). Much more than age. *Applied Linguistics*, *35*(4), 441–462. https://doi.org/10.1093/applin/amu032

Kirsner, K., Lalor, E., & Hird, K. (1993). The bilingual lexicon: Exercise, meaning and morphology. In R. Schreuder & B. Weltens (Eds.), *The bilingual lexicon* (pp. 215–248). John Benjamins Publishing Company.

Kirsner, K., & Speelman, C. (1996). Skill acquisition and repetition priming: One principle, many processes? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(3), 563–575. https://doi.org/10.1037/0278-7393.22.3.563

Klatt, D. H., & Stefanski, R. A. (1974). How does a mynah bird imitate human speech? *The Journal of the Acoustical Society of America*, *55*(4), 822–832. https://doi.org/10.1121/1.1914607

Krashen, S. (1977). The monitor model for adult second language performance. In M. Burt, H. Dulay, & M. Finocchiaro (Eds.), *Viewpoints on English as a second language* (pp. 152–161). Regents.

Krashen, S. (1981). *Second language acquisition and second language learning*. Pergamon.

Krashen, S. (1982). *Principles and practice in second language acquisition*. Pergamon.

Krashen, S. (1985). *The input hypothesis: Issues and implications*. Longman.

Krogh, S. M. (2022). Danish-English bilinguals' cognate processing in L1 and L2 visual lexical decision tasks. *Languages*, *7*, 228. https://doi.org/10.3390/languages7030228

Kroll, J. F., & Merves, J. S. (1986). Lexical access for concrete and abstract words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(1), 92–107. https://doi.org/10.1037/0278-7393.12.1.92

Kroll, J. F., Michael, E., & Sankaranarayan, A. (1998). A model of bilingual representation and its implications for second language acquisition. In A. F. Healy & L. E. Borne (Eds.), *Foreign language learning: Psycholinguistics studies on training and retention* (pp. 365–395). Lawrence Erlbaum Associates.

Kroll, J. F., Michael, E., Tokowicz, N., & Dufour, R. (2002). The development of lexical fluency in a second language. *Second Language Research*, *18*(2), 137–171. https://doi.org/10.1191/0267658302sr201oa

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, *33*(2), 149–174. https://doi.org/10.1006/jmla.1994.1008

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English*. Brown University Press.

Lado, R. (1957). *Linguistics across cultures: Applied linguistics for language teachers*. The University of Michigan Press.

Laka, I. (1996). *A brief grammar of Euskara, the Basque language*. Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU).

Lalor, E., & Kirsner, K. (2000). Cross-lingual transfer effects between English and Italian cognates and noncognates. *International Journal of Bilingualism*, *4*(3), 385–398. https://doi.org/10.1177/13670069000040030501

Lalor, E., & Kirsner, K. (2001). The representation of "false cognates" in the bilingual lexicon. *Psychonomic Bulletin & Review*, *8*(3), 552–559. https://doi.org/10.3758/BF03196191

Lee, E.-K., Lu, D. H.-Y., & Garnsey, S. M. (2013). L1 word order and sensitivity to verb bias in L2 processing. *Bilingualism: Language and Cognition*, *16*(4), 761–775. https://doi.org/10.1017/S1366728912000776

Leeman, J. (2003). Recasts and second language development: Beyond negative evidence. *Studies in Second Language Acquisition*, *25*(1), 37–63. https://doi.org/10.1017/S0272263103000020

Leeman, J. (2007). Feedback in L2 learning: Responding to errors during practice. In R. M. DeKeyser (Ed.), *Practice in a second language: Perspectives from linguistics and psychology* (pp. 111–137). Cambridge University Press.

Lehtonen, M., Hultén, A., Rodríguez-Fornells, A., Cunillera, T., Tuomainen, J., & Laine, M. (2012). Differences in word recognition between early bilinguals and monolinguals: Behavioral and ERP evidence. *Neuropsychologia*, *50*(7), 1362–1371. https://doi.org/10.1016/j.neuropsychologia.2012.02.021

Lemhöfer, K., & Dijkstra, T. (2004). Recognizing cognates and interlingual homographs: Effects of code similarity in language-specific and generalized lexical decision. *Memory & Cognition*, *32*(4), 533–550. https://doi.org/10.3758/BF03195845

Lemhöfer, K., Dijkstra, T., Schriefers, H., Baayen, R. H., Grainger, J., & Zwitserlood, P. (2008). Native language influences on word recognition in a second language: A megastudy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(1), 12–31. https://doi.org/10.1037/0278-7393.34.1.12

Leow, R. P. (2000). A study of the role of awareness in foreign language behavior: Aware versus unaware learners. *Studies in Second Language Acquisition*, *22*(4), 557–584. https://doi.org/10.1017/S0272263100004046

Leung, J., & Williams, J. N. (2006). Implicit learning of form-meaning connections. In R. Sun & N. Miyake (Eds.), *Proceedings of the 28th annual conference of the Cognitive Science Society* (pp. 465–470). Lawrence Erlbaum Associates.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(1), 1–38. https://doi.org/10.1017/S0140525X99001776

Levshina, N. (2015). *How to do linguistics with R: Data exploration and statistical analysis*. John Benjamins Publishing Company.

Lewicki, P. (1986). *Nonconscious social information processing*. Academic Press.

Li, C., & Wei, L. (2023). Anxiety, enjoyment, and boredom in language learning amongst junior secondary students in rural China: How do they contribute to L2 achievement? *Studies in Second Language Acquisition*, *45*(1), 93–108. https://doi.org/10.1017/S0272263122000031

Li, S., Hiver, P., & Papi, M. (Eds.). (2022). *The Routledge handbook of second language acquisition and individual differences*. Routledge.

Libben, M. R., & Titone, D. A. (2009). Bilingual lexical access in context: Evidence from eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*(2), 381–390. https://doi.org/10.1037/a0014875

Linck, J. A., & Weiss, D. J. (2011). Working memory predicts the acquisition of explicit L2 knowledge. In C. Sanz & R. P. Leow (Eds.), *Implicit and explicit language learning: Conditions, processes and knowledge in SLA and bilingualism* (pp. 101–113). Georgetown University Press.

Loebell, H., & Bock, K. (2003). Structural priming across languages. *Linguistics*, *41*(5), 791–824. https://doi.org/10.1515/ling.2003.026

Long, M. H., Inagaki, S., & Ortega, L. (1998). The role of implicit negative feedback in SLA: Models and recasts in Japanese and Spanish. *The Modern Language Journal*, *82*(3), 357–371. https://doi.org/10.1111/j.1540-4781.1998.tb01213.x

Longa, V. M. (1994). The Galician inflected infinitive and the theory of UG. *Catalan Working Papers in Linguistics*, *4*(1), 23–44.

López, B. (1997). Aportaciones de la tipología lingüística a una gramática particular: El concepto de orden básico y su aplicación al castellano. *Verba. Anuario Galego de Filoloxía*, *24*, 45–81.

Lotto, L., & de Groot, A. M. B. (1998). Effects of learning method and word type on acquiring vocabulary in an unfamiliar language. *Language Learning*, *48*(1), 31–69. https://doi.org/10.1111/1467-9922.00032

Luoni, G. (2022). *Developing a unified feature-based model for L2 lexical and syntactic processing* [Doctoral dissertation, Swansea University]. Cronfa-Swansea University's Research Repository. https://cronfa.swan.ac.uk/Record/cronfa59912

Luque, A., Mizyed, N., & Morgan-Short, K. (2018). Event-related potentials reveal evidence for syntactic co-activation in bilingual language processing. In L. López (Ed.), *Code-switching – Experimental answers to theoretical questions* (pp. 177–194). John Benjamins Publishing Company.

MacDonald, M. C., Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review*, *101*(4), 676–703. https://doi.org/10.1037/0033-295X.101.4.676

Mackey, A., & Philp, J. (1998). Conversational interaction and second language development: Recasts, responses, and red herrings? *The Modern Language Journal*, *82*(3), 338–356. https://doi.org/10.1111/j.1540-4781.1998.tb01211.x

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed). Lawrence Erlbaum Associates.

MacSwan, J. (2000). The architecture of the bilingual language faculty: Evidence from intrasentential code switching. *Bilingualism: Language and Cognition*, *3*(1), 37–54. https://doi.org/10.1017/S1366728900000122

MacWhinney, B. (1987a). The competition model. In B. MacWhinney (Ed.), *Mechanisms of language acquisition* (pp. 249–308). Lawrence Erlbaum Associates.

MacWhinney, B. (1987b). Applying the competition model to bilingualism. *Applied Psycholinguistics*, *8*(4), 315–327. https://doi.org/10.1017/S0142716400000357

MacWhinney, B. (1992). Transfer and competition in second language learning. In R. J. Harris (Ed.), *Advances in Psychology* (Vol. 83, pp. 371–390). Elsevier. https://doi.org/10.1016/S0166-4115(08)61506-X

MacWhinney, B. (1997). Second language acquisition and the competition model. In A. M. B. de Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 113–142). Lawrence Erlbaum Associates.

MacWhinney, B. (2005). A unified model of language acquisition. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 49–67). Oxford University Press.

Maie, R., & DeKeyser, R. M. (2020). Conflicting evidence of explicit and implicit knowledge from objective and subjective measures. *Studies in Second Language Acquisition*, *42*(2), 359–382. https://doi.org/10.1017/S0272263119000615

Makowski, D. (2018). The psycho package: An efficient and publishing-oriented workflow for psychological science. *The Journal of Open Source Software*, *3*(22), 470. https://doi.org/10.21105/joss.00470

Marecka, M., Szewczyk, J., Otwinowska, A., Durlik, J., Foryś-Nogala, M., Kutyłowska, K., & Wodniecka, Z. (2021). False friends or real friends? False cognates show advantage in word form learning. *Cognition*, *206*, 104477. https://doi.org/10.1016/j.cognition.2020.104477

Marian, V., & Spivey, M. (1999). Activation of Russian and English cohorts during bilingual spoken word recognition. In M. Hahn & S. C. Stoness (Eds.), *Proceedings of the 21st annual conference of the Cognitive Science Society* (pp. 349–354). Lawrence Erlbaum Associates.

Marian, V., & Spivey, M. (2003). Competing activation in bilingual language processing: Within- and between-language competition. *Bilingualism: Language and Cognition*, *6*(2), 97–115. https://doi.org/10.1017/S1366728903001068

Markina, E., & Garcia, A. (2022). The effect of a teacher-centred and learner-centred approach on students' participation in the English classroom. *Bellaterra Journal of Teaching & Learning Language & Literature*, *15*(3), e1007. https://doi.org/10.5565/rev/jtl3.1007

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1–2), 71–102. https://doi.org/10.1016/0010-0277(87)90005-9

Marslen-Wilson, W. D. (1990). Activation, competition, and frequency in lexical access. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 148–172). MIT Press.

McClelland, J. L. (1987). The case for interactionism in language processing. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 3–36). Lawrence Erlbaum Associates.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86. https://doi.org/10.1016/0010-0285(86)90015-0

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, *88*(5), 375–407. https://doi.org/10.1037/0033-295X.88.5.375

McClelland, J. L., St. John, M., & Taraban, R. (1989). Sentence comprehension: A parallel distributed processing approach. *Language and Cognitive Processes*, *4*(3–4), SI287–SI335. https://doi.org/10.1080/01690968908406371

McDonald, J. L. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language*, *55*(3), 381–401. https://doi.org/10.1016/j.jml.2006.06.006

Mcdonough, K., & Kim, Y. (2009). Syntactic priming, type frequency, and EFL learners' production of *wh-* questions. *The Modern Language Journal*, *93*(3), 386–398. https://doi.org/10.1111/j.1540-4781.2009.00897.x

McManus, K. (2015). L1-L2 differences in the acquisition of form-meaning pairings in a second language. *The Canadian Modern Language Review*, *71*(2), 155–181. https://doi.org/10.3138/cmlr.2070.51

McManus, K. (2022). *Crosslinguistic influence and second language learning*. Routledge.

McRae, K., & Matsuki, K. (2013). Constraint-based models of sentence processing. In R. P. G. van Gompel (Ed.), *Current issues in the psychology of language. Sentence processing* (pp. 51–77). Psychology Press.

Meng, M., & Bader, M. (2000). Mode of disambiguation and garden-path strength: An investigation of subject-object ambiguities in German. *Language and Speech*, *43*(1), 43–74. https://doi.org/10.1177/00238309000430010201

Miao, R. (2015). Second language acquisition: An introduction. In J. D. Wright (Ed.), *International encyclopedia of the social & behavioral sciences* (2nd ed., pp. 360–367). Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.92096-8

Midgley, K. J., Holcomb, P. J., & Grainger, J. (2011). Effects of cognate status on word comprehension in second language learners: An ERP investigation. *Journal of Cognitive Neuroscience*, *23*(7), 1634–1647. https://doi.org/10.1162/jocn.2010.21463

Mitchell, J., & Bowers, J. (2020). Priorless recurrent networks learn curiously. *Proceedings of the 28th International Conference on Computational Linguistics*, 5147–5158. https://doi.org/10.18653/v1/2020.coling-main.451

Monsell, S., Doyle, M. C., & Haggard, P. N. (1989). Effects of frequency on visual word recognition tasks: Where are they? *Journal of Experimental Psychology: General*, *118*(1), 43–71. https://doi.org/10.1037/0096-3445.118.1.43

Morgan-Short, K., Faretta-Stutenberg, M., Brill-Schuetz, K. A., Carpenter, H., & Wong, P. C. M. (2014). Declarative and procedural memory as individual differences in second language acquisition. *Bilingualism: Language and Cognition*, *17*(1), 56–72. https://doi.org/10.1017/S1366728912000715

Morgan-Short, K., Sanz, C., Steinhauer, K., & Ullman, M. T. (2010). Second language acquisition of gender agreement in explicit and implicit training conditions: An event-related potential study. *Language Learning*, *60*(1), 154–193. https://doi.org/10.1111/j.1467-9922.2009.00554.x

Morgan-Short, K., & Ullman, M. T. (2011). The neurocognition of second language. In S. M. Gass & A. Mackey (Eds.), *Handbook of second language acquisition* (pp. 282–299). Routledge.

Morton, J. (1969). Interaction of information in word recognition. *Psychological Review*, *76*(2), 165–178. https://doi.org/10.1037/h0027366

Mulder, K., Dijkstra, T., & Baayen, R. H. (2015). Cross-language activation of morphological relatives in cognates: The role of orthographic overlap and task-related processing. *Frontiers in Human Neuroscience*, *9*, 16. https://doi.org/10.3389/fnhum.2015.00016

Muñoz, C., & Singleton, D. (2007). Foreign accent in advanced learners: Two successful profiles. In L. Roberts, A. Gürel, S. Tatar, & L. Marti (Eds.), *EUROSLA Yearbook* (Vol. 7, pp. 171–190). John Benjamins Publishing Company. http://www.jbe-platform.com/content/journals/10.1075/eurosla.7.10mun

Muranoi, H. (2000). Focus on form through interaction enhancement: Integrating formal instruction into a communicative task in EFL classrooms. *Language Learning*, *50*(4), 617–673. https://doi.org/10.1111/0023-8333.00142

Murao, R., & Kajiro, T. (2017). The effects of syllable sequence frequency on EFL learners' speech recognition. *ARELE: Annual Review of English Language Education in Japan*, *28*, 161–175. https://doi.org/10.20581/arele.28.0_161

Muylle, M., Bernolet, S., & Hartsuiker, R. J. (2020). The role of case marking and word order in cross-linguistic structural priming in late L2 acquisition. *Language Learning*, *70*(S2), 194–220. https://doi.org/10.1111/lang.12400

Nakamura, D. (2012). Input skewedness, consistency, and order of frequent verbs in frequency-driven second language construction learning: A replication and extension of Casenhiser and Goldberg (2005) to adult second language acquisition. *International Review of Applied Linguistics in Language Teaching*, *50*(1), 1–37. https://doi.org/10.1515/iral-2012-0001

Norris, D. (1984). The effects of frequency, repetition and stimulus quality in visual word recognition. *The Quarterly Journal of Experimental Psychology*, *36*(3), 507–518. https://doi.org/10.1080/14640748408402174

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*, aac4716. https://doi.org/10.1126/science.aac4716

Otwinowska, A. (2016). *Cognate vocabulary in language acquisition and use: Attitudes, awareness, activation*. Multilingual Matters.

Paap, K. R., McDonald, J. E., Schvaneveldt, R. W., & Noel, R. W. (1987). Frequency and pronounceability in visually presented naming and lexical decision tasks. In M. Coltheart (Ed.), *Attention and performance XII: The psychology of reading* (pp. 221–243). Lawrence Erlbaum Associates.

Pastor, L. (2019). *Estrategias facilitadoras del procesamiento en lenguas OV-VO – Evidencia de corpus en euskera y castellano* [Doctoral dissertation, Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU)]. Archivo Digital Docencia Investigación. http://hdl.handle.net/10810/39966

Pastor, L. (2020). Sobre el orden básico de palabras en euskera. In I. Laka (Ed.), *Hitzak sarean. Pello Salabururi esker onez* (pp. 112–129). Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU).

Pauwels, A. (2014). Rethinking the learning of languages in the context of globalisation and hyperlingualism. In D. Abendroth-Timmer & E.-M. Hennig (Eds.), *Plurilingualism and multiliteracies: International research on identity construction in language education* (pp. 41–56). Peter Lang.

Peeters, D., Dijkstra, T., & Grainger, J. (2013). The representation and processing of identical cognates by late bilinguals: RT and ERP effects. *Journal of Memory and Language*, *68*(4), 315–332. https://doi.org/10.1016/j.jml.2012.12.003

Pepperberg, I. M. (1981). Functional vocalizations by an African grey parrot (Psittacus erithacus). *Zeitschrift Für Tierpsychologie*, *55*(2), 139–160. https://doi.org/10.1111/j.1439-0310.1981.tb01265.x

Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language*, *39*(4), 633–651. https://doi.org/10.1006/jmla.1998.2592

Pineda, A. (2021). The development of DOM in the diachrony of Catalan: (Dis)similarities with respect to Spanish. In J. Kabatek, P. Obrist, & A. Wall (Eds.), *Differential Object Marking in Romance* (pp. 243–278). De Gruyter. https://doi.org/10.1515/9783110716207-009

Pinker, S. (1994). *The language instinct: The new science of language and mind*. Penguin Press.

Pinker, S. (1997). *How the mind works*. Norton.

Pixelmator Team. (2022). *Pixelmator Pro* (2.3.7).

Platzack, C. (1996). The initial hypothesis of syntax: A minimalist perspective on language acquisition and attrition. In H. Clahsen (Ed.), *Generative perspectives on language acquisition: Empirical findings, theoretical considerations, crosslinguistic comparisons* (pp. 369–414). John Benjamins Publishing Company.

Poort, E. D., & Rodd, J. M. (2017). The cognate facilitation effect in bilingual lexical decision is influenced by stimulus list composition. *Acta Psychologica*, *180*, 52–63. https://doi.org/10.1016/j.actpsy.2017.08.008

Prather, J. F., Okanoya, K., & Bolhuis, J. J. (2017). Brains for birds and babies: Neural parallels between birdsong and speech acquisition. *Neuroscience & Biobehavioral Reviews*, *81, Part B*, 225–237. https://doi.org/10.1016/j.neubiorev.2016.12.035

Psychology Software Tools, Inc. (2016). *E-prime* (3.0). https://support.pstnet.com/

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Raposo, E. (1987). Case theory and Infl-to-Comp: The inflected infinitive in European Portuguese. *Linguistic Inquiry*, *18*(1), 85–109.

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, *14*(3), 191–201. https://doi.org/10.3758/BF03197692

Rayner, K., & Raney, G. E. (1996). Eye movement control in reading and visual search: Effects of word frequency. *Psychonomic Bulletin & Review*, *3*(2), 245–248. https://doi.org/10.3758/BF03212426

Rayner, K., Sereno, S. C., & Raney, G. E. (1996). Eye movement control in reading: A comparison of two types of models. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(5), 1188–1200. https://doi.org/10.1037/0096-1523.22.5.1188

Real Academia Española. (2021). *Corpus del Español del Siglo XXI (CORPES)*.
Real Academia Española, & Asociación de Academias de la Lengua Española. (2009). *Nueva gramática de la lengua española*. Espasa Libros.
Real Academia Galega, & Instituto da Lingua Galega. (2012). *Normas ortográficas e morfolóxicas do idioma galego* (23[a] ed). Real Academia Galega. https://publicacions.academia.gal/index.php/rag/catalog/book/252
Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, *6*(6), 855–863. https://doi.org/10.1016/S0022-5371(67)80149-X
Reber, A. S. (1969). Transfer of syntactic structure in synthetic languages. *Journal of Experimental Psychology*, *81*(1), 115–119. https://doi.org/10.1037/h0027454
Rebuschat, P. (2009). *Implicit learning of natural language syntax* [Doctoral dissertation, University of Cambridge]. Apollo-Repository of the University of Cambridge. https://doi.org/10.17863/CAM.15883
Rebuschat, P. (2013). Measuring implicit and explicit knowledge in second language research. *Language Learning*, *63*(3), 595–626. https://doi.org/10.1111/lang.12010
Rebuschat, P. (Ed.). (2015). *Implicit and explicit learning of languages*. John Benjamins Publishing Company.
Rebuschat, P., & Williams, J. N. (2012). Implicit and explicit knowledge in second language acquisition. *Applied Psycholinguistics*, *33*(4), 829–856. https://doi.org/10.1017/S0142716411000580
Rieder-Bünemann, A. (2012). Second language learning. In N. M. Seel (Ed.), *Encyclopedia of the sciences of learning* (pp. 2980–2983). Springer. https://doi.org/10.1007/978-1-4419-1428-6_826
Ringbom, H. (2007). *Cross-linguistic similarity in foreign language learning*. Multilingual Matters.
Robinson, P. (1996). Learning simple and complex second language rules under implicit, incidental, rule-search, and instructed conditions. *Studies in Second Language Acquisition*, *18*(1), 27–67. https://doi.org/10.1017/S0272263100014674
Robinson, P. (1997). Individual differences and the fundamental similarity of implicit and explicit adult second language learning. *Language Learning*, *47*(1), 45–99. https://doi.org/10.1111/0023-8333.21997002
Rogers, J., Révész, A., & Rebuschat, P. (2015). Challenges in implicit learning research: Validating a novel artificial language. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 275–299). John Benjamins Publishing Company.
Rogers, J., Webb, S., & Nakata, T. (2015). Do the cognacy characteristics of loanwords make them more easily learned than noncognates? *Language Teaching Research*, *19*(1), 9–27. https://doi.org/10.1177/1362168814541752
Rohrmeier, M., Rebuschat, P., & Cross, I. (2011). Incidental and online learning of melodic structure. *Consciousness and Cognition*, *20*(2), 214–222. https://doi.org/10.1016/j.concog.2010.07.004
Rosa, E. (1999). *A cognitive approach to task-based research: Explicitness, awareness and L2 development* [Doctoral dissertation, Georgetown University].

Rosa, E., & O'Neill, M. D. (1999). Explicitness, intake, and the issue of awareness: Another piece to the puzzle. *Studies in Second Language Acquisition*, *21*(4), 511–556. https://doi.org/10.1017/S0272263199004015

Rossi, S., Gugler, M. F., Friederici, A. D., & Hahne, A. (2006). The impact of proficiency on syntactic second-language processing of German and Italian: Evidence from event-related potentials. *Journal of Cognitive Neuroscience*, *18*(12), 2030–2048. https://doi.org/10.1162/jocn.2006.18.12.2030

Rugg, M. D. (1990). Event-related brain potentials dissociate repetition effects of high-and low-frequency words. *Memory & Cognition*, *18*(4), 367–379. https://doi.org/10.3758/BF03197126

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: II. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review*, *89*(1), 60–94. https://doi.org/10.1037/0033-295X.89.1.60

Sabourin, L., & Stowe, L. A. (2008). Second language processing: When are first and second languages processed similarly? *Second Language Research*, *24*(3), 397–430. https://doi.org/10.1177/0267658308090186

Sabourin, L., Stowe, L. A., & de Haan, G. J. (2006). Transfer effects in learning a second language grammatical gender system. *Second Language Research*, *22*(1), 1–29. https://doi.org/10.1191/0267658306sr259oa

Sagarra, N. (2017). Longitudinal effects of working memory on L2 grammar and reading abilities. *Second Language Research*, *33*(3), 341–363. https://doi.org/10.1177/0267658317690577

Sánchez, A. (2007). A learner-centred approach to the teaching of English as an L2. *ES: Revista de filología inglesa*, *28*, 189–196.

Sánchez-Casas, R. M., & García-Albea, J. E. (2005). The representation of cognate and noncognate words in bilingual memory: Can cognate status be characterized as a special kind of morphological relation? In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 226–250). Oxford University Press.

Sánchez-Casas, R. M., García-Albea, J. E., & Davis, C. W. (1992). Bilingual lexical processing: Exploring the cognate/non-cognate distinction. *European Journal of Cognitive Psychology*, *4*(4), 293–310. https://doi.org/10.1080/09541449208406189

Sanoudaki, E., & Thierry, G. (2015). Language non-selective syntactic activation in early bilinguals: The role of verbal fluency. *International Journal of Bilingual Education and Bilingualism*, *18*(5), 548–560. https://doi.org/10.1080/13670050.2015.1027143

Sarasola, I., Salaburu, P., Landa, J., & Zabaleta, J. (2009). *Ereduzko Prosa Gaur (EPG)*. Universidad del País Vasco/Euskal Herriko Unibertsitatea (UPV/EHU). http://www.ehu.eus/euskara-orria/euskara/ereduzkoa/

Scarborough, D. L., Cortese, C., & Scarborough, H. S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance*, *3*(1), 1–17. https://doi.org/10.1037/0096-1523.3.1.1

Schepens, J., Dijkstra, T., & Grootjen, F. (2012). Distributions of cognates in Europe as based on Levenshtein distance. *Bilingualism: Language and Cognition*, *15*(1), 157–166. https://doi.org/10.1017/S1366728910000623

Schepens, J., van der Slik, F., & van Hout, R. (2013). The effect of linguistic distance across Indo-European mother tongues on learning Dutch as a second language. In L. Borin & A. Saxena (Eds.), *Approaches to measuring linguistic differences* (pp. 199–233). De Gruyter Mouton.

Schilling, H. E. H., Rayner, K., & Chumbley, J. I. (1998). Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory & Cognition*, *26*(6), 1270–1281. https://doi.org/10.3758/BF03201199

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*(2), 129–158. https://doi.org/10.1093/applin/11.2.129

Schmidt, R. (1992). Awareness and second language acquisition. *Annual Review of Applied Linguistics*, *13*, 206–226. https://doi.org/10.1017/S0267190500002476

Schmitt, N. (1997). Vocabulary learning strategies. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary. Description, acquisition and pedagogy* (pp. 199–227). Cambridge University Press.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-prime* (2.0). Psychology Software Tools, Inc.

Schoonbaert, S., Hartsuiker, R. J., & Pickering, M. J. (2007). The representation of lexical and syntactic information in bilinguals: Evidence from syntactic priming. *Journal of Memory and Language*, *56*(2), 153–171. https://doi.org/10.1016/j.jml.2006.10.002

Schulpen, B., Dijkstra, T., Schriefers, H. J., & Hasper, M. (2003). Recognition of interlingual homophones in bilingual auditory word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(6), 1155–1178. https://doi.org/10.1037/0096-1523.29.6.1155

Schwartz, A. I., & Kroll, J. F. (2006). Bilingual lexical activation in sentence context. *Journal of Memory and Language*, *55*(2), 197–212. https://doi.org/10.1016/j.jml.2006.03.004

Schwartz, B. D. (1999). Let's make up your mind: "Special nativist" perspectives on language, modularity of mind, and nonnative language acquisition. *Studies in Second Language Acquisition*, *21*(4), 635–655. https://doi.org/10.1017/S0272263199004052

Schwartz, B. D., & Sprouse, R. A. (1994). Word order and nominative case in non-native language acquisition: A longitudinal study of (L1 Turkish) German interlanguage. In T. Hoekstra & B. D. Schwartz (Eds.), *Language acquisition and language disorders* (Vol. 8, pp. 317–368). John Benjamins Publishing Company. https://doi.org/10.1075/lald.8.14sch

Schwartz, B. D., & Sprouse, R. A. (1996). L2 cognitive states and the full transfer/full access model. *Second Language Research*, *12*(1), 40–72. https://doi.org/10.1177/026765839601200103

Schwartz, B. D., & Sprouse, R. A. (2021). In defense of 'copying and restructuring'. *Second Language Research*, *37*(3), 489–493. https://doi.org/10.1177/0267658320975831

Sharwood Smith, M. (1983). Cross-linguistic aspects of second language acquisition. *Applied Linguistics*, *4*(3), 192–199. https://doi.org/10.1093/applin/4.3.192

Sharwood Smith, M. (2017). *Introducing language and cognition: A map of the mind*. Cambridge University Press.

Sharwood Smith, M., & Truscott, J. (2006). Full Transfer Full Access: A processing-oriented interpretation. In S. Unsworth, T. Parodi, A. Sorace, & M. Young-Scholten (Eds.), *Paths of development in L1 and L2 acquisition* (pp. 201–216). John Benjamins Publishing Company.

Sharwood Smith, M., & Truscott, J. (2014). *The multilingual mind: A modular processing perspective*. Cambridge University Press.

Sheehan, M., & Parafita Couto, M. C. (2011). *Optional inflected infinitives in Portuguese and Galician [Paper presentation]*. Poznań Linguistics Meeting, Poznań, Poland.

Shin, J.-A., & Christianson, K. (2009). Syntactic processing in Korean–English bilingual production: Evidence from cross-linguistic structural priming. *Cognition*, *112*(1), 175–180. https://doi.org/10.1016/j.cognition.2009.03.011

Skinner, B. F. (1957). *Verbal behavior.* Appleton-Century-Crofts. https://doi.org/10.1037/11256-000

Smith, N. V., & Tsimpli, I.-M. (1995). *The mind of a savant: Language learning and modularity*. Blackwell.

Soares, A. P., Fraga, I., Comesaña, M., & Piñeiro, A. (2010). The role of animacy in European Portuguese relative clause attachment: Evidence from production and comprehension tasks. *Psicothema*, *22*(4), 691–696.

Soares, A. P., Oliveira, H., Ferreira, M., Comesaña, M., Macedo, A. F., Ferré, P., Acuña-Fariña, C., Hernández-Cabrera, J., & Fraga, I. (2019). Lexico-syntactic interactions during the processing of temporally ambiguous L2 relative clauses: An eye-tracking study with intermediate and advanced Portuguese-English bilinguals. *PLOS ONE*, *14*(5), e0216779. https://doi.org/10.1371/journal.pone.0216779

Soares, A. P., Oliveira, H. M., Comesaña, M., & Costa, A. S. (2018). Lexico-syntactic interactions in the resolution of relative clause ambiguities in a second language (L2): The role of cognate status and L2 proficiency. *Psicológica Journal*, *39*(2), 164–197. https://doi.org/10.2478/psicolj-2018-0008

Sonderegger, M., Wagner, M., & Torreira, F. (2018). *Quantitative methods for linguistic data*. McGill University. http://people.linguistics.mcgill.ca/~morgan/qmld-book/

Speelman, C. P., & Kirsner, K. (2005). *Beyond the learning curve: The construction of mind*. Oxford University Press.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*(1), 137–149. https://doi.org/10.3758/BF03207704

Staub, A. (2011). Word recognition and syntactic attachment in reading: Evidence for a staged architecture. *Journal of Experimental Psychology: General*, *140*(3), 407–433. https://doi.org/10.1037/a0023517

Strijkers, K., Costa, A., & Thierry, G. (2010). Tracking lexical access in speech production: Electrophysiological correlates of word frequency and cognate effects. *Cerebral Cortex*, *20*(4), 912–928. https://doi.org/10.1093/cercor/bhp153

Sunderman, G., & Schwartz, A. I. (2008). Using cognates to investigate cross-language competition in second language processing. *TESOL Quarterly*, *42*(3), 527–536. https://doi.org/10.1002/j.1545-7249.2008.tb00145.x

Suzuki, Y., & DeKeyser, R. M. (2017). The interface of explicit and implicit knowledge in a second language: Insights from individual differences in cognitive aptitudes. *Language Learning*, *67*(4), 747–790. https://doi.org/10.1111/lang.12241

Tagarelli, K. M., Borges-Mota, M., & Rebuschat, P. (2011). The role of working memory in implicit and explicit language learning. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the*

33rd annual conference of the Cognitive Science Society (pp. 2061–2066). Cognitive Science Society.

Tagarelli, K. M., Ruiz, S., Vega, J. L. M., & Rebuschat, P. (2016). Variability in second language learning: The roles of individual differences, learning conditions, and linguistic complexity. *Studies in Second Language Acquisition*, *38*(2), 293–316. https://doi.org/10.1017/S0272263116000036

Thierry, G., & Sanoudaki, E. (2012). Activation syntaxique non-sélective à la langue chez le bilingue précoce. *Revue Française de Linguistique Appliquée*, *17*(2), 33–48. https://doi.org/10.3917/rfla.172.0033

Thomas, J. (1988). The role played by metalinguistic awareness in second and third language learning. *Journal of Multilingual and Multicultural Development*, *9*(3), 235–246. https://doi.org/10.1080/01434632.1988.9994334

Tily, H., Fedorenko, E., & Gibson, E. (2010). The time-course of lexical and structural processes in sentence comprehension. *Quarterly Journal of Experimental Psychology*, *63*(5), 910–927. https://doi.org/10.1080/17470210903114866

Tokowicz, N., Kroll, J. F., de Groot, A. M. B., & van Hell, J. G. (2002). Number-of-translation norms for Dutch—English translation pairs: A new tool for examining language production. *Behavior Research Methods, Instruments, & Computers*, *34*(3), 435–451. https://doi.org/10.3758/BF03195472

Tokowicz, N., & MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar: An event-related potential investigation. *Studies in Second Language Acquisition*, *27*(2), 173–204. https://doi.org/10.1017/S0272263105050102

Tolentino, L. C., & Tokowicz, N. (2014). Cross-language similarity modulates effectiveness of second language grammar instruction: L2 instruction and cross-language similarity. *Language Learning*, *64*(2), 279–309. https://doi.org/10.1111/lang.12048

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.

Tomasello, M. (2007). If they're so good at grammar, then why don't they talk? Hints from apes' and humans' use of gestures. *Language Learning and Development*, *3*(2), 133–156. https://doi.org/10.1080/15475440701225451

Tonzar, C., Lotto, L., & Job, R. (2009). L2 vocabulary acquisition in children: Effects of learning method and cognate status. *Language Learning*, *59*(3), 623–646. https://doi.org/10.1111/j.1467-9922.2009.00519.x

Traxler, M. J., Morris, R. K., & Seely, R. E. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of Memory and Language*, *47*(1), 69–90. https://doi.org/10.1006/jmla.2001.2836

Traxler, M. J., & Pickering, M. J. (1996). Plausibility and the processing of unbounded dependencies: An eye-tracking study. *Journal of Memory and Language*, *35*(3), 454–475. https://doi.org/10.1006/jmla.1996.0025

Traxler, M. J., Williams, R. S., Blozis, S. A., & Morris, R. K. (2005). Working memory, animacy, and verb class in the processing of relative clauses. *Journal of Memory and Language*, *53*(2), 204–224. https://doi.org/10.1016/j.jml.2005.02.010

Trueswell, J. C. (1996). The role of lexical frequency in syntactic ambiguity resolution. *Journal of Memory and Language*, *35*(4), 566–585. https://doi.org/10.1006/jmla.1996.0030

Trueswell, J. C., & Tanenhaus, M. K. (1994). Toward a lexicalist framework for constraint-based syntactic ambiguity resolution. In C. Jr. Clifton, L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 155–179). Lawrence Erlbaum Associates.

Trueswell, J. C., Tanenhaus, M. K., & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*(3), 528–553. https://doi.org/10.1037/0278-7393.19.3.528

Ullman, M. T. (2001a). A neurocognitive perspective on language: The declarative/procedural model. *Nature Reviews Neuroscience*, *2*(10), 717–726. https://doi.org/10.1038/35094573

Ullman, M. T. (2001b). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition*, *4*(2), 105–122. https://doi.org/10.1017/S1366728901000220

Ullman, M. T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition*, *92*(1–2), 231–270. https://doi.org/10.1016/j.cognition.2003.10.008

Ullman, M. T. (2008). The role of memory systems in disorders of language. In B. Stemmer & H. A. Whitaker (Eds.), *Handbook of the neuroscience of language* (pp. 189–198). Elsevier.

Ullman, M. T. (2012). The declarative/procedural model. In P. Robinson (Ed.), *Routledge encyclopedia of second language acquisition* (pp. 160–164). Routledge.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498–505. https://doi.org/10.3758/BF03192720

Vafaee, P., Suzuki, Y., & Kachisnke, I. (2017). Validating grammaticality judgment tests: Evidence from two new psycholinguistic measures. *Studies in Second Language Acquisition*, *39*(1), 59–95. https://doi.org/10.1017/S0272263115000455

Vainikka, A., & Young-Scholten, M. (1994). Direct access to X'-theory: Evidence from Korean and Turkish adults learning German. In T. Hoekstra & B. D. Schwartz (Eds.), *Language acquisition and language disorders* (Vol. 8, pp. 265–316). John Benjamins Publishing Company. https://doi.org/10.1075/lald.8.13vai

Vainikka, A., & Young-Scholten, M. (1996a). Gradual development of L2 phrase structure. *Second Language Research*, *12*(1), 7–39. https://doi.org/10.1177/026765839601200102

Vainikka, A., & Young-Scholten, M. (1996b). The early stages in adult L2 syntax: Additional evidence from Romance speakers. *Second Language Research*, *12*(2), 140–176. https://doi.org/10.1177/026765839601200202

Valente, D., Ferré, P., Soares, A. P., Rato, A., & Comesaña, M. (2018). Does phonological overlap of cognate words modulate cognate acquisition and processing in developing and skilled readers? *Language Acquisition*, *25*(4), 438–453. https://doi.org/10.1080/10489223.2017.1395029

Van Assche, E., Drieghe, D., Duyck, W., Welvaert, M., & Hartsuiker, R. J. (2011). The influence of semantic constraints on bilingual word recognition during sentence reading. *Journal of Memory and Language*, *64*(1), 88–107. https://doi.org/10.1016/j.jml.2010.08.006

Van Assche, E., Duyck, W., & Brysbaert, M. (2013). Verb processing by bilinguals in sentence contexts: The effect of cognate status and verb tense. *Studies in Second Language Acquisition*, *35*(2), 237–259. https://doi.org/10.1017/S0272263112000873

Van Assche, E., Duyck, W., Hartsuiker, R. J., & Diependaele, K. (2009). Does bilingualism change native-language reading? : Cognate effects in a sentence context. *Psychological Science*, *20*(8), 923–927. https://doi.org/10.1111/j.1467-9280.2009.02389.x

van Gompel, R. P. G., & Pickering, M. J. (2007). Syntactic parsing. In M. G. Gaskill (Ed.), *The Oxford handbook of psycholinguistics.* (pp. 289–307). Oxford University Press.

van Hell, J. G., & Dijkstra, T. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic Bulletin & Review*, *9*(4), 780–789. https://doi.org/10.3758/BF03196335

van Heuven, W. J. B., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language*, *39*(3), 458–483. https://doi.org/10.1006/jmla.1998.2584

Van Orden, G. C. (1987). A ROWS is a ROSE: Spelling, sound, and reading. *Memory & Cognition*, *15*(3), 181–198. https://doi.org/10.3758/BF03197716

Van Valin, R. D. (2001). *An introduction to syntax*. Cambridge University Press.

Van Wijnendaele, I., & Brysbaert, M. (2002). Visual word recognition in bilinguals: Phonological priming from the second to the first language. *Journal of Experimental Psychology: Human Perception and Performance*, *28*(3), 616–627. https://doi.org/10.1037/0096-1523.28.3.616

Vanlangendonck, F., Peeters, D., Rueschemeyer, S.-A., & Dijkstra, T. (2020). Mixing the stimulus list in bilingual lexical decision turns cognate facilitation effects into mirrored inhibition effects. *Bilingualism: Language and Cognition*, *23*(4), 836–844. https://doi.org/10.1017/S1366728919000531

VanPatten, B. (1996). *Input processing and grammar instruction in second language acquisition*. Ablex Publishing Corporation.

VanPatten, B. (2002). Processing instruction: An update. *Language Learning*, *52*(4), 755–803. https://doi.org/10.1111/1467-9922.00203

VanPatten, B. (2004). Input processing in second language acquisition. In B. VanPatten (Ed.), *Processing instruction: Theory, research, and commentary* (pp. 5–31). Lawrence Erlbaum Associates.

VanPatten, B., & Cadierno, T. (1993). Explicit instruction and input processing. *Studies in Second Language Acquisition*, *15*(2), 225–243. https://doi.org/10.1017/S0272263100011979

VanPatten, B., & Oikkenon, S. (1996). Explanation versus structured input in processing instruction. *Studies in Second Language Acquisition*, *18*(4), 495–510. https://doi.org/10.1017/S0272263100015394

VanPatten, B., & Williams, J. (Eds.). (2015). *Theories in second language acquisition: An introduction*. Routledge.

Vaughan-Evans, A., Liversedge, S. P., Fitzsimmons, G., & Jones, M. W. (2020). Syntactic co-activation in natural reading. *Visual Cognition*, *28*(10), 541–556. https://doi.org/10.1080/13506285.2020.1841866

Voga, M., & Grainger, J. (2007). Cognate status and cross-script translation priming. *Memory & Cognition*, *35*(5), 938–952. https://doi.org/10.3758/BF03193467

Wang, A.-L. (2023). *Understanding bilingualism, bilinguality, and bilingual education in an era of globalization*. IGI Global. https://doi.org/10.4018/978-1-6684-4869-4

Weber-Fox, C. M., & Neville, H. J. (1996). Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. *Journal of Cognitive Neuroscience*, *8*(3), 231–256. https://doi.org/10.1162/jocn.1996.8.3.231

Westergaard, M. (2021). Microvariation in multilingual situations: The importance of property-by-property acquisition. *Second Language Research*, *37*(3), 379–407. https://doi.org/10.1177/0267658319884116

Whaley, C. P. (1978). Word-nonword classification time. *Journal of Verbal Learning and Verbal Behavior*, *17*(2), 143–154. https://doi.org/10.1016/S0022-5371(78)90110-X

White, L. (2003). *Second language acquisition and Universal Grammar*. Cambridge University Press.

Whitford, V., & Titone, D. (2012). Second-language experience modulates first- and second-language word frequency effects: Evidence from eye movement measures of natural paragraph reading. *Psychonomic Bulletin & Review*, *19*(1), 73–80. https://doi.org/10.3758/s13423-011-0179-5

Whitley, M. S. (2002). *Spanish-English contrasts: A course in Spanish linguistics* (2nd ed). Georgetown University Press.

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. Springer.

Williams, J. N. (2005). Learning without awareness. *Studies in Second Language Acquisition*, *27*(2), 269–304. https://doi.org/10.1017/S0272263105050138

Williams, J. N. (2006). Incremental interpretation in second language sentence processing. *Bilingualism: Language and Cognition*, *9*(1), 71–88. https://doi.org/10.1017/S1366728905002385

Williams, J. N., & Kuribara, C. (2008). Comparing a nativist and emergentist approach to the initial stage of SLA: An investigation of Japanese scrambling. *Lingua*, *118*(4), 522–553. https://doi.org/10.1016/j.lingua.2007.03.003

Williams, J. N., & Lovatt, P. (2003). Phonological memory and rule learning. *Language Learning*, *53*(1), 67–121. https://doi.org/10.1111/1467-9922.00211

Winter, B. (2020). *Statistics for linguists: An introduction using R*. Routledge.

Xiong, K., Verdonschot, R. G., & Tamaoka, K. (2020). The time course of brain activity in reading identical cognates: An ERP study of Chinese - Japanese bilinguals. *Journal of Neurolinguistics*, *55*, 100911. https://doi.org/10.1016/j.jneuroling.2020.100911

Year, J., & Gordon, P. (2009). Korean speakers' acquisition of the English ditransitive construction: The role of verb prototype, input distribution, and frequency. *The Modern Language Journal*, *93*(3), 399–417. https://doi.org/10.1111/j.1540-4781.2009.00898.x

Yu, L., & Odlin, T. (Eds.). (2016). *New perspectives on transfer in second language learning*. Multilingual Matters.

Zawiszewski, A., Gutiérrez, E., Fernández, B., & Laka, I. (2011). Language distance and non-native syntactic processing: Evidence from event-related potentials. *Bilingualism: Language and Cognition*, *14*(3), 400–411. https://doi.org/10.1017/S1366728910000350

Zawiszewski, A., Martinez De La Hidalga, G., & Laka, I. (2022). Agents strongly preferred: ERP evidence from natives and non-natives processing intransitive sentences in Spanish. *Brain Sciences*, *12*(7), 853. https://doi.org/10.3390/brainsci12070853

Ziegler, J. C., Perry, C., Jacobs, A. M., & Braun, M. (2001). Identical words are read differently in different languages. *Psychological Science*, *12*(5), 379–384. https://doi.org/10.1111/1467-9280.00370