Journal of Memory and Language 128 (2023) 104386

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

Interference between non-native languages during trilingual language production

Angela de Bruin[a,*], Liv J. Hoversten[b], Clara D. Martin[c,d]

[a] Department of Psychology, University of York, York, UK
[b] Department of Psychology, University of California, Santa Cruz, USA
[c] Basque Center on Cognition, Brain and Language (BCBL), Donostia-San Sebastián, Spain
[d] Ikerbasque, Basque Foundation for Science, Bilbao, Spain

ARTICLE INFO

Keywords:
Trilingualism
Language production
Language interference
Inhibition

ABSTRACT

Most research on multilingual language control has focused on a bilingual's first (L1) and second (L2) languages. Studies on third language (L3) acquisition suggest that, despite the L1 being more proficient, L3 learners experience more L2 than L1 interference. However, little is known about how a trilingual's L2 and L3 interact after initial stages of language learning. In the current study (Experiment 1: 30 Spanish-Basque-English trilinguals; Experiment 2: 50 English-French-Spanish trilinguals), participants completed a speeded naming task to assess cross-language intrusions (e.g., using the Spanish "perro" instead of the French "chien"). Both experiments showed more L3 than L1 intrusions during L2 naming. Furthermore, using two different tasks, we assessed if this cross-language interference was related to language inhibition. Both experiments suggested that trilinguals inhibited their L1 more strongly than their L3. Together, this suggests that a trilingual's non-native language might experience more interference from another non-native language than from their L1, possibly because trilinguals apply more inhibition over their L1.

Introduction

While most research on multilingualism focuses on a bilingual's first (L1) and second language (L2), a large proportion of the world population can hold a conversation in more than two languages (e.g., European Commission, Special Eurobarometer 386, 2012). Studies looking at third language (L3) acquisition suggest that while learning a new language, there might be more interference from the other non-native language (L2) than from the native language (L1), despite the L1 being more proficient (e.g., Puig-Mayenco, González Alonso, and Rothman, 2020). However, it remains largely unknown how the languages of a trilingual influence each other once fluency is reached in all three languages (i.e., after initial stages of language learning). Across two experiments, this study therefore investigated how a trilingual's native and non-native languages interact and compete with each other during language production.

Interactions between non-native languages

Research looking at interactions between native and multiple non-native languages has mainly focused on people acquiring a new language (L3). Despite the L1 having a higher proficiency than the L2, several studies have suggested L3 acquirers might be influenced more (e.g., using syntactic structures from a known language in the L3) by the L2 than L1 (e.g., Bardel & Falk, 2007; Falk & Bardel, 2011; Rothman & Cabrelli Amaro, 2010). In a recent systematic review, Puig-Mayenco et al. (2020) examined the influence from the L1 versus L2 on third language acquisition across 71 studies, focusing on morphosyntax. Twenty of the reviewed studies showed influence exclusively from the L2 while this was the case for the L1 in only ten studies. While there are many variables that could have an impact on L3 acquisition (with typological proximity being one of the key factors), this suggests that L3 acquisition might experience more influence from another non-native language (L2) than from the native L1.

While morphosyntax has been the focus of much of this research, there is also some experimental evidence showing that competition between non-native languages is also present at the lexical level. Mickan, McQueen, and Lemhöfer (2020) asked Dutch-English bilinguals to learn a set of Spanish words. The next day, participants were asked to name the same pictures in either Dutch or in English. Participants then

* Corresponding autor at: Department of Psychology, University of York, York YO10 5DD, UK.
E-mail address: angela.debruin@york.ac.uk (A. de Bruin).

https://doi.org/10.1016/j.jml.2022.104386
Received 11 April 2022; Received in revised form 26 October 2022; Accepted 28 October 2022
Available online 9 November 2022

completed another naming task to assess Spanish recall immediately after the Dutch/English naming task and one week later. Spanish naming (in terms of accuracy and reaction times) was affected negatively by Dutch and English naming. However, this influence was largest for the words that had been previously named in English (L2). This suggests that at the lexical level too, a non-native language might experience more influence from another non-native language than from the native language.

One explanation for these findings during L3 acquisition has been given in the form of the "L2 status" hypothesis (e.g., Bardel & Falk, 2007; 2012), which argues that the L2 and L3 are more cognitively similar than the L1 and L2/L3. Trilinguals might acquire the L2 and L3 in similar environments (e.g., in a classroom as opposed to at home or in the community) and in similar life stages (e.g., in later childhood or adulthood as opposed to from birth). The L1 and L2/L3 might also differ in their reliance on procedural versus declarative memory processes, with the latter potentially being more important for languages acquired later in life (e.g., Bardel & Sánchez, 2017). Together, these similarities in the way non-native languages are acquired (and potentially used, if language use is restricted to e.g., the classroom) could explain why the L2 might influence L3 acquisition more than the L1.

Most of this research has focused on L3 acquisition, however, and leaves open the question how the languages of a trilingual interact once a certain level of fluency is achieved in all languages. There is some anecdotal evidence to suggest that interference continues to exist between the non-native languages beyond initial stages of acquisition. For example, Williams and Hammarberg (1998) report a case study with an English-German-Swedish trilingual who, when switching out of their L3 Swedish, switched more often to their L2 German than to their L1 English. More recently, Tomoschuk, Duyck, Hartsuiker, Ferreira, and Gollan (2021) provided experimental evidence showing similar patterns, with more interference between non-native languages than between a native and non-native language. In their Experiment 1, they asked Dutch-English-French trilinguals (with a high proficiency in their L1 and L2 but a lower proficiency in their L3) to complete a phoneme detection task. Participants had to indicate whether a phoneme was present in the word corresponding to a picture (e.g., /g/ when presented with a picture of a "girl"). Importantly, participants were also presented with phonemes that were not part of the word in the target language but that were part of the translation equivalent (e.g., in the case of a picture of a girl, the phoneme /m/ is present in the L1 translation equivalent "meisje"). If there is more interference from a native language, L3 blocks should see more false alarms from the L1 (i.e., saying that the phoneme is present in the target word because it is present in the L1 translation equivalent) than from the L2. However, if there is more interference between non-native languages, there should be more L2 than L1 false alarms. The latter pattern was found, with more false alarms from L2 than L1 phonemes during the L3 task, suggesting that the L3 experienced more interference from the L2 than from the native L1.

Language competition and inhibition

Tomoschuk et al. (2021) suggest that the increased interference from the L2 might be related to inhibition, with trilinguals inhibiting their L1 more strongly or successfully than their L2. This interpretation is based on Green's Inhibitory Control Hypothesis (1998), which posits that bilinguals use inhibition to avoid interference from the non-target language(s) to allow for successful production in the target language. Importantly, this hypothesis argues that the amount of inhibition applied is relative to the proficiency of the language, with bi-/trilinguals suppressing more proficient languages more strongly than less proficient languages.

Evidence that bilinguals might inhibit their L1 (more so than their L2) has been found in a range of paradigms and techniques. Language switching studies (e.g., Meuter & Allport, 1999) have suggested that unbalanced bilinguals (with a higher proficiency in their L1 than L2) can show asymmetrical switch costs, with larger costs when switching back to their L1 than L2. This might be the consequence of bilinguals applying relatively large amounts of L1 inhibition during L2 naming, thus requiring more time to release this inhibition when switching back to the L1, leading to larger switching costs. The finding that more inhibition is applied over the L1 when switching to the L2 than vice versa is also supported by neuroimaging studies showing increased activation in brain regions and ERP markers associated with inhibition when switching to an L2 (i.e., the moment L1 inhibition would need to be applied) than when switching to an L1 (i.e., when less (L2) inhibition is needed; de Bruin, Roelofs, Dijkstra, & FitzPatrick, 2014; Jackson, Swainson, Cunnington, & Jackson, 2001). Other studies not using language-switching tasks have also suggested that L1 words are less accessible after L2 naming due to L1 inhibition during L2 naming. For example, Misra, Guo, Bobb, and Kroll (2012) assessed potential effects of repetition priming (i.e., faster naming when items have been named before) when naming in the L2 after the L1 and in the L1 after the L2. As expected, repetition priming was observed when the L2 was used after the L1. However, using the L1 second showed no behavioural repetition benefit and an increased N2 (an ERP component associated with top-down language control), suggesting that the L1 was suppressed during L2 naming.

Interference and inhibition have also been studied through paradigms eliciting language intrusions. For example, Gollan, Schotter, Gomez, Murillo, and Rayner (2014) asked Spanish-English bilinguals to read paragraphs in one language or mixing both languages aloud. In the texts with both Spanish and English words, participants made cross-language intrusions (e.g., saying an English word instead of the Spanish word), in particular when trying to read words in the more dominant language. This again suggests that the more dominant/proficient language might be suppressed and consequently less accessible and more prone to intrusions. Other paradigms too have suggested that cross-language intrusions during production are sensitive to language control and interference (e.g., Declerck, Grainger, & Hartsuiker, 2021; Zheng, Roelofs, & Lemhöfer, 2020), although they do not always reveal differences between languages (e.g., Declerck et al., 2021).

Levy, McVeigh, Marful, & Anderson (2007) also examined the accessibility of L1 versus L2 word forms, by using a retrieval-induced forgetting paradigm, a task we used in Experiment 1 too and which we will refer to as the "rhyme task". In Levy et al.'s study, English-Spanish bilinguals were asked to generate English rhyme words. Participants saw a probe (e.g., "spoon") and had to generate one word that rhymed with the probe (e.g., "moon"). Prior to the rhyme task, participants named pictures in English or in Spanish. The names of those pictures could be used as rhyme words in the rhyme task (e.g., participants named a picture of a moon, with "spoon" being the probe in the rhyme task). Naming those pictures in English increased the chance of using those words in the English rhyme task (i.e., picture names produced in English were more likely to be used as rhyme responses if those words had been repeated more often in the picture-naming task). In contrast, the reverse pattern was found for words produced in Spanish (L2): the more often a picture had been named in Spanish prior to the rhyme task, the less likely participants were to use the English (L1) translation equivalent in the rhyme task. This suggests that participants had suppressed the L1 equivalents during L2 naming, thus making those L1 translation equivalents less accessible in the following rhyme task (but cf. Runnqvist & Costa, 2012 for diverging findings).

Current research

Following these findings, unbalanced trilinguals might suppress their more proficient native language more strongly than their less proficient non-native languages. This, in turn, could be predicted to reduce interference from a native language. This prediction is also supported by studies suggesting that L3 learning is more successful through L1 instruction than through L2 instruction, potentially because of better

Fig. 1. Overview of the tasks in each session in Experiment 1.

regulation and inhibition of the L1 (e.g., Bogulski, Bice, & Kroll, 2019; Hirosh & Degani, 2021). However, while there is some evidence to suggest that trilinguals might experience more interference between non-native languages than from a native language (Tomoschuk et al., 2021), most research is based on L3 *acquisition*. Furthermore, Tomoschuk et al. (2021) showed interference in a phoneme detection task but it is unknown how potential interference between non-native languages can actually influence language production. Lastly, while L1 inhibition has been proposed as a potential underlying mechanism to explain reduced L1 interference relative to L2/L3 interference (Tomoschuk et al., 2021), the role of language inhibition has not been assessed directly and we do not know if and how language inhibition explains potential interference between languages during trilingual production.

The current research therefore firstly aimed to examine how the two non-native languages of a trilingual interact with each other beyond initial stages of language acquisition, when trilinguals have already achieved an intermediate proficiency in their non-native languages. Specifically, we addressed interference between languages in the form of cross-language intrusions, such as saying the Spanish word "caballo" when the Basque "zaldi" is needed. Second, we examined *why* trilinguals might experience interference between non-native languages by assessing the role of L1 versus L3 inhibition. Across two experiments with trilinguals with different language combinations, we therefore used a speeded picture-naming task to study interference **from the L1 versus L3 on the L2** during trilingual language production. We also examined how trilinguals **inhibited their L1 versus L3** during L2 production through a rhyme task (Experiment 1) and an n-2 switching task (Experiment 2).

Experiment 1

Introduction

To assess language interference, Experiment 1 asked participants to complete a language-switching task in three languages (Spanish-L1, Basque-L2, and English-L3). We were especially interested in eliciting language intrusions (e.g., using a Spanish or English word when a Basque word was intended). Cross-language intrusions can be a more direct and concrete measure of language interference as they reveal noticeable mistakes in trilingual language selection. As such, this allowed us to study which language (L1 or L3) interfered more with L2 production. We therefore developed a speeded picture-naming paradigm in which participants had to alternate languages in a task presenting each picture for less than one second. If trilinguals experience more interference between non-native languages, despite a higher proficiency level in their native language, we hypothesised that they should produce more L3 than L1 intrusions during L2 target naming.

As a second aim, we wanted to examine the role of inhibition during trilingual language control. To this end, we adjusted the retrieval-induced forgetting paradigm (rhyme task) used by Levy et al. (2007). In this task, participants generated a rhyme word in response to a probe (e.g., "sheep" in response to the probe "jeep"). They completed this task in Spanish (L1) and in English (L3) before and after a trilingual naming task (a different naming task than the speeded switching task). During the trilingual naming task, they named some pictures in their L1, some in their L2, and some in their L3. During the rhyme task, the probe rhymed with words used in the naming task (within-language trials) or their translation equivalents (across-language trials, see Fig. 1 and the Methods section for further information). For example, participants could see the English probe "cake" in the rhyme task pre- and post-naming. During the naming task, they would be asked to name pictures of a snake in English, thus increasing the likelihood of participants saying "snake" in the post-naming rhyme task (compared to the baseline pre-naming rhyme task). These so-called "within-language" control trials were used for both English-L3 and Spanish-L1 (e.g., in Spanish, a rhyme probe could be "rama", with participants using "cama" to name the picture of a bed in the naming task). We expected within-language targets (i.e., "snake" and "cama") to be used more often in the rhyme task post- than pre-naming due to their recent use in the naming task. However, of main interest, we also included "across-language" targets. These were L1/L3 response targets to the rhyme probe that were translation equivalents of words that had just been named in the L2 in the naming task. For example, participants would need to name a picture of a glass in Basque ("edalontzi") in the naming task and would see the probe "mass" in English (rhyming with "glass") and "paso" in Spanish (rhyming with "vaso", meaning glass). In the pre-naming rhyme task we did not expect differences between languages given that the words had not yet been used in the naming task. However, in the post-naming

Table 1
Summary of the objective and subjective measures of language proficiency, language exposure, and language use for Spanish, Basque, and English.

	Spanish			Basque			English		
	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range
Age of Acquisition	0	0	0–0	2.9	0.8	2–6	6.1	2.0	2–11
Picture naming (0–65)	64.8	0.5	64–65	42.6	6.7	30–51	43.8	8.0	31–56
LexTALE (% words correct)[1]	88.7	8.0	83–100	67.4	16.4	28–94	43.1	22.2	3–100
Interview (1–5)[2]	5	0	5–5	3.3	0.4	3–4	3.0	0.6	2–4
Self-rated proficiency[2] (0–10)									
Speaking	9.6	0.7x	7–10	6.9	1.5	4–10	6.0	1.7	1–9
Understanding	9.7	0.7	8–10	8.3	1.3	6–10	7.0	1.5	3–9
Writing	9.3	1.0	7–10	7.3	1.7	4–10	6.3	1.9	2–9
Reading	9.4	1.2	5–10	7.9	1.7	5–10	7.0	1.7	3–9
General	9.4	0.8	8–10	7.0	1.4	4–10	6.3	1.6	2–9
%exposure (0–100)	64.3	13.3	50–90	24.6	14.3	0–40	10.4	7.4	0–30
%speaking (0–100)	74.3	16.5	50–100	19.3	14.6	0–40	15.2	12.4	0–40

[1] Data from one participant missing for Basque and English. LexTALE score is calculated as the percentage of words identified correctly.
[2] Data are missing for two participants.

rhyme task we expected language differences on the across-language targets. Specifically, if trilinguals inhibit L1 equivalents more than L3 equivalents during L2 naming, we would expect them to use those L1 equivalents less often than the L3 equivalents during the post-naming rhyme task. For example, if a trilingual names a picture of a glass in L2 ("edalontzi") and concurrently suppresses L1 "vaso" more than L3 "glass", we would expect them to use "vaso" less often in response to the L1 rhyme probe "paso" than they would use "glass" to the L3 probe "mass".

Experiment 1 took place in the Basque Country. Participants had a lower proficiency in their L2/L3 than L1 and acquired only their L1 from birth. However, their language environment differs from the type of trilinguals/L3 acquirers tested in previous studies. For example, in many other studies (e.g., Tomoschuk et al., 2021), the non-native languages were both acquired as classroom languages. According to the L2 status hypothesis, interference between non-native languages could be explained by cognitive similarity as a consequence of the L2 and L3 being acquired in a classroom. In the Basque Country, however, the participants' L2 (Basque) is a community language while the participants' L3 (English) is treated as a foreign language that is largely restricted to the classroom. If the trilinguals in Experiment 1 show more L3 than L1 intrusions, this would suggest that this interference is not purely due to the way the L2 and L3 were acquired and are used in similar (classroom) environments.

Furthermore, most participants received their (primary and secondary) education in Basque (L2) or a dual Basque-Spanish system. Tomoschuk et al. (2021, Experiment 2) suggested that language of instruction can modulate the amount of non-native language interference. Interference differences between the L1 and L2 were only found if new L3 words were taught in the L1, but not when the instruction language was L2. This raises the question whether trilinguals indeed experience more interference from their non-native language *in general* or whether these findings are specific to the L1 being used as the language of instruction in the classroom. If trilinguals in our Experiment 1 show more L3 than L1 intrusions (despite the L3 being taught in an L2 or bilingual school environment), this would suggest that they are better at regulating interference from their native language even if the L1 was not the main/only language of instruction.

In addition, although their L2 Basque had a much lower proficiency and use than their L1 Spanish, most participants acquired Basque within the first three years of life, making it less likely that the L2 was acquired mostly through declarative memory processes. According to the L2 status hypothesis (e.g., Bardel & Sánchez, 2017), reliance on more declarative than procedural memory systems for the L2 and L3 might explain the interference between the two. However, due to the way Basque is acquired during early childhood and is a community language while English is largely used as a classroom language, the L2 and L3 in this study are unlikely to be both relying predominantly on declarative memory.

Finally, typological proximity has been argued to be a key variable in language interference, with more interactions between languages that are more similar (e.g., Puig-Mayenco et al, 2020). Basque (L2) differs substantially from both the L1 and L3 in aspects such as vocabulary and morphosyntax. However, if anything, it is more similar to the L1 than L3 in terms of orthography and phonology. Thus, if our trilingual participants experience more interference from the L3 than L1, this is unlikely to be due to typological proximity between the L2 and L3.

Taking into consideration this different language profile compared to previous L3 (acquisition) studies, Experiment 1 firstly aimed to assess how much interference trilinguals experience between two non-native languages as compared to between a native and non-native language using the speeded naming task. To create the largest difference between the native and non-native language (in terms of proficiency, use, mode of acquisition, etcetera), we compared L1 versus L3 intrusions during L2 production. Next, we assessed how inhibition might be involved when trilinguals manage interference between their three languages using the rhyme task.

Data availability

The data and analysis script (for both Experiments) are available on https://osf.io/wmehd/. The stimuli are provided in the appendices.

Methods

Participants

The final dataset included thirty Spanish-Basque-English trilinguals (23 female, *M*age = 23.3; *SD*age = 5.6). We specifically recruited participants who acquired only one language from birth (Spanish) and who had an intermediate proficiency level in Basque and English. We ensured that participants were proficient in all languages (thus focusing on trilinguals rather than L3 acquirers) but had a much lower use of and proficiency in their L2/L3 than their L1 (thus creating a clear distinction in proficiency and use between the native L1 and non-native L2 and L3). Two additional participants were tested but excluded from data analysis. One participant was excluded because they did not know over half of the English target words (assessed in a post-experiment survey). The other participant was excluded because they did not produce any rhyme words on more than half of the trials in the baseline rhyme task. The use of novel tasks (a speeded switching task to elicit intrusions and a trilingual

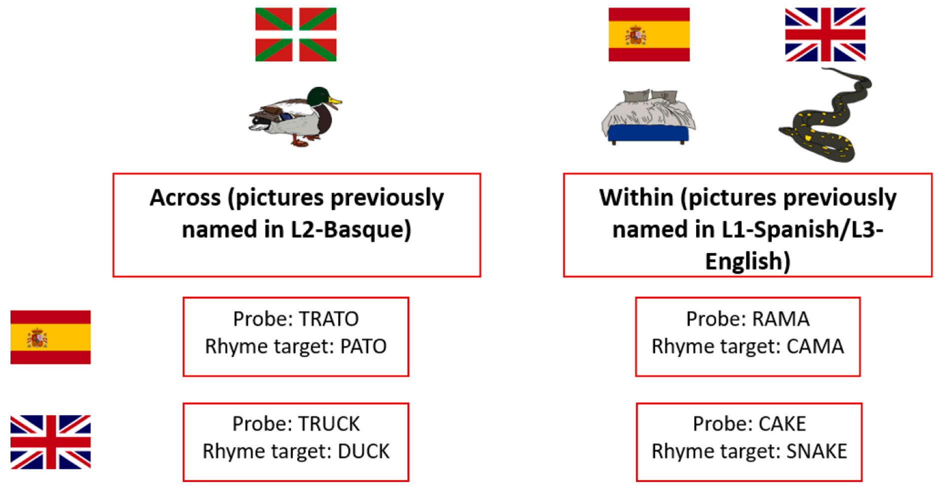

Fig. 2. Overview of the conditions in the rhyme task.

adaptation of a rhyme task) meant that we did not have existing data on which to base a power analysis. Furthermore, the specific language profile restricted us in the number of participants we could recruit. Thirty participants was therefore set as the target size to ensure recruitment feasibility while aiming to have as much power as possible. All participants had normal or corrected-to-normal vision and no known neurological, reading, or hearing impairments. They provided written informed consent and the study was approved by the BCBL Ethics Review Board and complied with the guidelines of the Helsinki Declaration.

All participants acquired Spanish as their first language, Basque as their second language, and English as their third language. They were living in a bilingual society in which both Spanish and Basque are commonly used. Participants had an intermediate proficiency in both Basque and English and predominantly used Spanish on a daily basis (see Table 1). Their language profile was assessed through a set of objective and subjective language measures in Spanish, Basque, and English when they signed up for the database (de Bruin, Carreiras, & Duñabeitia, 2017). The objective proficiency measures include a 65-item picture naming task, the LexTALE (a short lexical decision task; Lemhöfer & Broersma, 2012) and an interview. In addition, participants provided self-ratings of language proficiency, use, and exposure. The results from these tasks and measures are reported in Table 1. Participants also completed a short survey at the end of the study asking about their language(s) of education and time spent in English-speaking countries. Most participants (18) attended a bilingual Spanish-Basque school system; 10 participants completed their education in Basque; two participants completed their education in Spanish. Most participants (19) had not spent any significant time in an English-speaking country apart from holidays. Eleven participants indicated having spent some time in an English-speaking country (*M* number of months = 4.2, *SD* = 4.2).

Design of the two main tasks

Speeded trilingual naming task. In the speeded trilingual naming task, participants were asked to name pictures in Spanish (L1), Basque (L2), or English (L3) in response to a country flag. A speeded task (in which participants saw each picture for only 900 ms) was used to create more time pressure and thus to induce more errors. We focused on the number of Spanish (L1) versus English (L3) intrusions during non-switch trials that were supposed to be named in Basque (L2).

Rhyme task. In the rhyme task, participants were asked to generate rhyme words in response to Spanish (L1) and English (L3) probes. This task was completed twice: once at the beginning of the study (as the baseline, we will refer to this as "pre-naming") and once at the end of the study after a naming task ("post-naming"; see Fig. 1). In the intervening naming task (a different one than the speeded naming task) before the rhyme task, people named words in either L1, L2, or L3. Words named in the L1 and L3 in that task were possible rhyme responses (i.e., targets) in the L1/L3 rhyme task (e.g., "snake" had to be named in English in the naming task and "cake" was a probe in the rhyme task). We refer to this condition as "within-language". The words named in L2 were translation equivalents of possible rhyme responses in the L1/L3 rhyme task (e.g., "duck" named in Basque in the naming task; probe "truck"; see Fig. 2). We refer to this condition as "across-language". There were thus three within-subject independent variables in the rhyme task: Language (L1/L3); Session (pre-naming/post-naming); Condition (words named in the same language "within-language"/words named in Basque "across-language").

Session 1:

Materials

We selected sixty pictures from the MultiPic database (Duñabeitia et al., 2018). Eight of these pictures were used in the speeded trilingual naming task and had to be named in the three languages interchangeably. We used a small set of pictures in this task to increase activation of each word, with the overall aim of increasing competition between words and eliciting more intrusions.

All sixty pictures were used in the slow naming task preceding the post-naming rhyme task (twenty named in Spanish serving as L1 within-language targets in the rhyme task; twenty named in English serving as L3 within-language targets in the rhyme task, and twenty named in Basque serving as both L1 and L3 across-language targets in the rhyme task). All picture names were non-cognates and had a high frequency.

The eight words used in the speeded naming task were matched on English and Spanish frequency and on number of letters and phonemes; Spanish words had more syllables than English words (see Appendix A). We did not match the Basque words given that we were focusing on the number of English (L3) versus Spanish (L1) intrusions but did match the English and Spanish words on their Levenshtein distance to Basque words (i.e., the number of letters that would need to be changed to form the Basque word). Those eight words were chosen (from the across-language items) because they differed in their onset in the three languages. This allowed us to score cross-language intrusions even if only one phoneme was produced (e.g., if the initial response was corrected after the first phoneme).

For the rhyme task, we matched the different conditions (within-

English versus within-Spanish; across-English versus across-Spanish; within-English versus across-English; within-Spanish versus across-Spanish) on a range of measures including target frequency, the number of potential rhyme words for each probe, and the number of alternative rhyme words with a higher frequency than the target (see Appendix A). Due to general word length differences between English and Spanish words, the English and Spanish words could not be matched in terms of number of syllables, letters, and phonemes. However, we did ensure that the Spanish-within and Spanish-across as well as the English-within and English-across conditions were matched on word length. We selected target words that had a few high-frequency rhyme competitors but not too many (to make sure that target words were relatively likely to be produced without the target being the only option). Further details about the probes are given in Appendix A.

Procedure

The study consisted of two in-person sessions (see Fig. 1), separated by an interval of approximately one to two weeks ($M = 12$ days). We opted to use two separate sessions to separate the baseline and post-naming rhyme tasks (to avoid participants remembering exactly which answers they gave in the baseline task) and to separate the two naming tasks. In the first session, participants started with the baseline pre-naming rhyme condition. Next, they completed the speeded trilingual naming task assessing cross-language intrusions during L2 production. In the second session, participants first completed the slow trilingual naming task that was related to the rhyme task. Next, they completed the post-naming rhyme task. After the second session, participants were asked to indicate whether there were any English words they did not know before the start of the study. On average, participants indicated knowing 38 of the 40 English rhyme targets (range 36–40; data missing from three participants).

Pre-naming rhyme task (baseline)

In the rhyme task, participants were presented visually and aurally with a probe word. They were asked to generate a word that rhymed with the probe. For example, a participant could be presented with 'wig' and could respond with 'pig', 'dig', etc. Participants were instructed to say the first rhyme word that came to mind, but not to use any proper names or names of countries or places. They were given ten seconds for each probe word. The task was completed in Spanish and English in two separate language blocks, with the order of languages counterbalanced across participants. Each language included twenty within-language trials and twenty across-language trials (see "slow naming task"). The rhyme task was completed at the start of the first session as the baseline for the rhyme task at the end of the second session.

Speeded trilingual naming task

Participants were first familiarised with the 8 pictures and words and were asked to read each word aloud. Participants were then presented with a speeded naming task showing pictures accompanied by the Spanish, Basque, or British flag. Each picture was presented below the country flag for 900 ms and participants were instructed to name the picture in the indicated language within that time frame. A blank screen was shown for one second before the next picture was presented. We recorded the full 1900 ms and scored responses that were given during the entire interval, even though participants were instructed to name the picture while the picture remained on the screen.

In total, participants named 720 experimental trials (240 Basque non-switch trials, 160 Basque switch trials, 40 English non-switch trials, 120 English switch trials, 40 Spanish non-switch trials, and 120 Spanish switch trials). Participants named an additional ten trials that were preceded by a break and were not included in the analysis. We focused on the 240 Basque non-switch trials to examine the number of L1 (Spanish) versus L3 (English) intrusions. Crucially, these L2 non-switch trials were always preceded by another L2 trial, thus removing the immediate influence of just having used another language. L2 switch trials were preceded an equal number of times by each of the languages (e.g., 80 Basque switch trials were preceded by a Spanish trial and 80 Basque switch trials were preceded by an English trial). L1 and L3 switch trials were preceded 80 times by a Basque trial and 40 times by an L3 or L1 trial respectively.

Session 2:

Slow naming task:

The slow naming task at the start of the second session was used to put rhyme targets in the "within-language" or "across-language" condition for the post-naming rhyme task. Before starting the naming task, participants were exposed to all sixty pictures and their names in the three languages to ensure participants were familiar with the responses. In the naming task, each picture had to be named eight times and was shown on the screen for 2 s below the country flag. Twenty of these pictures always had to be named in Basque (L2 – across-language trials for the L1 and L3 rhyme task), twenty always had to be named in English (L3 – within-language trials for the L3 rhyme task), and twenty always had to be named in Spanish (L1 – within-language trials for the L1 rhyme task). The three languages had to be used interchangeably to increase competition between the languages. There were 488 trials (8 trials preceded by a break; the 480 trials were distributed equally across languages and switch type). Each picture was presented four times as a switch trial (twice preceded by each of the two languages) and four times as a non-switch trial.

Post-naming rhyme task:

The rhyme task from the start of the first session was completed again, using the same probes and structure.

Analysis

The data are available on https://osf.io/wmehd/ and were analysed using generalised linear mixed-effects models using lme4 package version 1.1–21 in R 3.6.1.

Speeded trilingual naming task. In the speeded trilingual naming task, we scored accuracy on each trial as A) no response, B) a cross-language intrusion (e.g., English instead of Basque word), C) a within-language intrusion (e.g., 'cloud' instead of 'moon'), or D) another response that was not the intended target but that had a similar meaning (e.g., 'pared' instead of 'muro', with both being Spanish words for 'wall'). We only scored the first response. For example, if multiple cross-language intrusions were made (e.g., English and then Spanish where a Basque word was required), the first was scored. Similarly, a response counted as an intrusion regardless of how much of the intrusion was produced (e.g., if just the first phoneme of the intrusion was produced and then corrected, it still counted as an intrusion).

We were mainly interested in the number of L1 Spanish versus L3 English intrusions produced during L2 Basque non-switch trials. Our main analysis therefore only included L2 non-switch trials. Using generalised linear mixed-effect models (glmer; participant and item intercepts included as random effects), we compared the number of L1 versus L3 intrusions by using number of cross-language intrusions (i.e., accuracy type B) as the dependent variable (DV) and the language of intrusion (Spanish = -0.5; English = 0.5) as the fixed effect. We used the "poisson" distribution within the glmer, which is suitable for count data (number of intrusions in Spanish or English) and is based on the underlying data being dichotomous (an intrusion happening or not). We also examined the same question for L2 switch trials, now also including

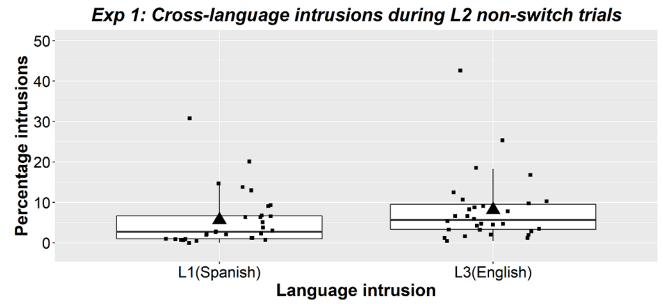

Fig. 3. Boxplots showing the percentage of language intrusions (L1 versus L3) during L2 non-switch trials. Each black square shows an individual participant (jittered). The horizontal line shows the median while the black triangle shows the mean.

the language of the previous trial (i.e., L1-L2 or L3-L2 switch) as a fixed effect (switching from L3 = 0.5; switching from L1 = -0.5).

Rhyme task. In the rhyme task, we scored whether the response to the rhyme probe was the target word (i.e., the word used in the naming task) or not (i.e., a different word was produced or there was no response). The DV used was Target response (0 = no, 1 = yes). The three independent variables (IVs) were Language (L1 −0.5; L3 0.5); Session (pre-naming −0.5; post-naming 0.5); and Condition (across language −0.5; within language 0.5). The model converged with intercepts for participants and items and the by-participant slope for condition.

Results

Speeded trilingual naming task

Across the 720 trials in the speeded trilingual naming task (excluding the ten trials after the breaks), participants on average answered 70.5 % correctly (*SD* = 13.3, range = 37–91 % correct). Participants did not respond on 13.9 % of trials (*SD* = 6.9). Cross-language intrusions (across all languages) were made on 14.2 % of all trials (*SD* = 12.2) and within-language intrusions on 1.1 % of trials (*SD* = 0.8); 0.3 % of trials (*SD* = 0.6) were both cross- and within-language intrusions (for example saying “apple” in Spanish when “horse” had to be named in English). On 0.6 % of trials (*SD* = 0.9), a correct response was given that was not the intended target word (e.g., “muro” instead of “pared”).

All participants produced cross-language intrusions, the error type of interest. We focused on the number of cross-language intrusions during Basque (L2) non-switch trials (see Fig. 3). During these trials, participants produced significantly more L3 (*M* = 8.2 % of L2 non-switch trials showed an L3 intrusion, *SD* = 8.6 %) than L1 (*M* = 5.7 % of L2 non-switch trials showed an L1 intrusion, *SD* = 6.9; β = 0.366, *SE* = 0.064, *z* = 5.719, $p < 0.001$) intrusions.

Basque (L2) switch trials showed similar findings. Participants produced more L3 (*M* = 9.2 %, *SD* = 8.6) than L1 (*M* = 6.5 %, *SD* = 6.8) intrusions (β = 0.344, *SE* = 0.073, *z* = 4.695, $p < 0.001$). There was no significant effect of the language used on the previous trial (β = 0.012, *SE* = 0.073, *z* = 0.160, *p* = 0.873) nor a significant interaction between language of previous trial and language of intrusion (β = 0.267, *SE* = 0.147, *z* = 1.818, *p* = 0.069). Participants not only made more L3 (*M* = 9.9 %, *SD* = 8.4) than L1 (*M* = 6.1 %, *SD* = 7.1) intrusions when an L2 trial was preceded by an L3 trial but also when it was preceded by an L1 trial (L3 intrusions *M* = 8.5 %, *SD* = 9.3; L1 intrusions *M* = 6.9 %, *SD* = 7.1).

Given the lower number of trials to be named in L3 (160 across trial types) or in L1 (160 across trial types), and given our focus on L2 production, we did not further analyse the number of intrusions during L1 or L3 trials. However, numerically more L3 intrusions were made during L1 trials (*M* = 5.3 %, *SD* = 7.2) than L1 intrusions during L3 trials (*M* = 4.0 %, *SD* = 6.9). Given that the task required participants to use the L2 more than the other languages, the largest number of intrusions during L1/L3 trials came from the L2 (during L1 trials *M* = 8.2 %, *SD* = 5.7; during L3 trials *M* = 9.7 %, *SD* = 7.2).

In summary, the speeded naming task shows that trilinguals were more likely to make L3 than L1 intrusions when having to name pictures in L2.

Rhyme task

The rhyme task included Session (pre-/post-naming), Language (L1/L3), and Condition (across-/within-language naming) as the variables.

Although the naming task separating the pre- and post-naming rhyme tasks was not of main interest, we examined accuracy to make sure participants named the pictures correctly. Accuracy was high in all three languages (Spanish *M* = 96.6 %, *SD* = 4.0; Basque *M* = 89.5 %, *SD* = 7.7; English *M* = 89.9 %, *SD* = 8.3). The majority of errors (*M* = 6.2 % of all trials; *SD* = 4.3) were no or late responses; cross-language intrusions (*M* = 0.6 % of trials, *SD* = 0.9) and non-target word choice (e.g., “pared” instead of “muro”, *M* = 1.2 % of trials, *SD* = 1.5) were rare. This confirms that participants used the target words in the naming task as intended.

Fig. 4 shows the results from the rhyme task. Participants produced significantly more rhyme targets in the post-naming task (*M* = 36.4 %, *SD* = 7.5) than in the baseline pre-naming rhyme task (*M* = 20.0 %, *SD* = 5.5; β = 1.052, *SE* = 0.078, *z* = 13.561, $p < 0.001$). There were no main effects of language (β = 0.145, *SE* = 0.307, *z* = 0.472, *p* = 0.637) or condition (β = 0.306, *SE* = 0.313, *z* = 0.978, *p* = 0.328). These two variables did not interact with each other (β = -0.841, *SE* = 0.615, *z* = -1.368, *p* = 0.171) but importantly they did interact with session. The interaction between session and language (β = 0.701, *SE* = 0.155, *z* = 4.537, $p < 0.001$) reflected that the increase between pre- and post-naming was larger for L3 (English) rhyme responses (*Mpre* = 18.3 %, *SD* = 6.0; *Mpost* = 40.3 %, *SD* = 9.7) than for L1 (Spanish) rhyme responses (*Mpre* = 21.8 %, *SD* = 7.2; *Mpost* = 32.5 %, *SD* = 7.4). Session also interacted with condition (β = 0.421, *SE* = 0.155, *z* = 2.723, *p* =

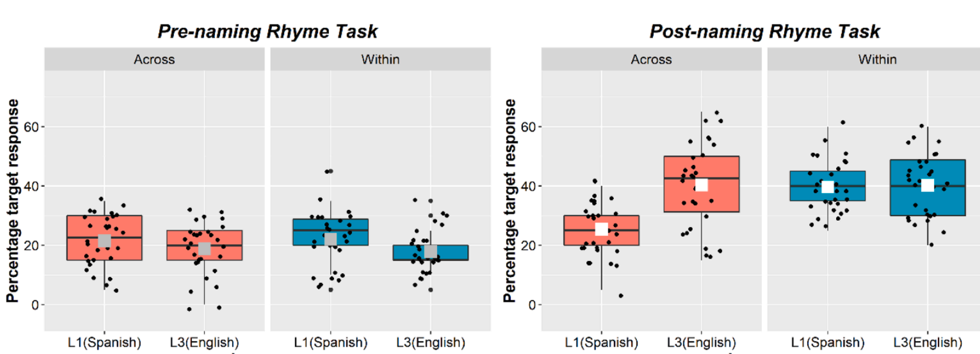

Fig. 4. Boxplots showing the percentage of target rhyme responses in the pre-naming task (left, baseline) and the post-naming task (right). Within each plot, the left panel represents the ‘across-language’ condition (targets named in L2 Basque) and the right panel the ‘within-language’ condition (targets named in Spanish/English, i.e., in the same language as the rhyme task). Each black dot shows an individual participant. The horizontal line shows the median while the white square shows the mean.

0.006), reflecting that the increase between pre- and post-naming was larger for within-language trials (i.e., trials previously named in the same language as the rhyme task; $Mpre = 19.9$ %, $SD = 6.3$; $Mpost = 39.9$ %, $SD = 8.5$) than for across-language trials (i.e., trials previously named in L2; $Mpre = 20.2$ %, $SD = 7.3$; $Mpost = 32.8$ %, $SD = 9.7$). Importantly, there was a three-way interaction between language, session, and condition ($\beta = -0.744$, $SE = 0.309$, $z = -2.408$, $p = 0.016$). As Fig. 4 shows, there was a comparable increase relative to the baseline in target rhyme responses for L1 and L3 within-language trials. However, there was a larger increase compared to baseline in L3 across-language than L1 across-language trials, suggesting that L1 across-language targets were less accessible than L3 across-language targets.

Follow-up analyses for the pre- and post-task separately first showed that during the pre-naming rhyme task (i.e., baseline), crucially, there was no interaction between language and condition ($\beta = -0.230$, $SE = 0.653$, $z = -0.353$, $p = 0.724$). This was expected given that the across-/within-language condition was only introduced in the naming task and was therefore irrelevant in the pre-naming baseline. The post-naming rhyme task, however, did show an interaction between language and condition ($\beta = -1.208$, $SE = 0.611$, $z = -1.978$, $p = 0.048$). While there was a similar number of L3 across- and within-language rhyme targets ($\beta = -0.089$, $SE = 0.412$, $z = -0.215$, $p = 0.830$), there were more L1 within- than across-language rhyme targets ($\beta = 1.119$, $SE = 0.453$, $z = 2.467$, $p = 0.014$). This suggests that while L3 equivalents previously named in L2 were not less accessible than those named in the L3 itself, L1 translation equivalents previously named in L2 were less accessible than words named in the L1 itself.

Follow-up analyses for the two conditions separately first showed an increase in target responses post- compared to pre-naming in the within-language condition ($\beta = 1.247$, $SE = 0.107$, $z = 11.659$, $p < 0.001$). Crucially, this did not interact with language ($\beta = 0.325$, $SE = 0.213$, $z = 1.527$, $p = 0.127$), demonstrating that the increase in target responses after using those words in the L1/L3 naming task did not differ for the L1 and L3. In contrast, the across-language trials showed an interaction between session and language ($\beta = 1.090$, $SE = 0.225$, $z = 4.847$, $p < 0.001$), reflecting a larger increase in across-language targets for the L3 than L1. In the L3, there was a significant increase in across-language rhyme targets between the pre- and post-naming task ($\beta = 1.395$, $SE = 0.156$, $z = 8.922$, $p < 0.001$). In the L1 there was only a small increase in across-language rhyme targets that did not reach significance ($\beta = 0.310$, $SE = 0.162$, $z = 1.915$, $p = 0.056$).

To summarise the findings from the rhyme task, all conditions showed a similar number of target responses in the baseline measure. After the slow naming task, there was a significant increase in target responses. For words previously named in the actual language (within-language), this increase did not differ for the L1 and L3. However, for words previously named in the L2 (across-language), this increase was larger for the L3 than L1, suggesting L1 translation equivalents were less accessible than L3 equivalents.

Correlation between intrusions and rhyme targets

We also assessed whether there was a correlation between the relative number of L1 intrusions during L2 non-switch trials (compared to L3 intrusions) and the number of L1 across-language post-naming rhyme words produced (relative to L3). In other words, we aimed to assess here whether the measure of inhibition (rhyme task) was related to the intrusions made. We were specifically interested in assessing whether people who made *relatively* few L1 intrusions (as compared to L3 intrusions) also produced *relatively* few L1 rhyme targets (as compared to L3 targets). To do this, we used the percentage difference between L1 and L3 intrusions and the percentage difference in rhyme target words between L1 and L3 in the across-language condition post-naming. There was no significant correlation between the intrusions and rhyme targets ($r = -0.156$, $p = 0.410$; see Supplementary Fig. 1). As an additional post-hoc check, we also examined the correlation while taking into consideration performance during the baseline rhyme task (to exclude any individual differences in how likely participants were to use target words before the naming manipulation). The correlation was not observed either when we computed the L1-L3 difference on the post-naming rhyme task relative to the pre-naming rhyme task ($r = -0.225$, $p = 0.232$; see Supplementary Fig. 1). In both cases, the direction of the correlation was the opposite of expected (although not significant). This is largely driven by one outlier (see Supplementary Fig. 1) who produced a very large number of L3 intrusions.

Discussion

Experiment 1 had two aims. First, we examined the amount of language interference (in the form of cross-language intrusions) stemming from the native language (L1) versus from a weaker non-native language (L3) during L2 production. During a speeded picture-naming task participants showed more L3 than L1 intrusions during L2 production. This suggests that, despite the L1 being far more proficient, there was more interference from the non-native language than from the native language. Second, we aimed to examine whether trilinguals suppressed the L1 more strongly than the L3 during L2 production. Using a rhyme task after a picture-naming task showed that L1 translation equivalents were used less often than L3 translation equivalents, possibly because they were suppressed more strongly during L2 naming and consequently less accessible.

The larger amount of L3 than L1 intrusions during L2 use is consistent with the L3 acquisition literature (e.g., Bardel & Falk, 2007; Falk & Bardel, 2011; Mickan et al., 2020; Puig-Mayenco et al., 2020; Rothman & Cabrelli Amaro, 2010), which shows that a non-native language has a larger influence than the native language while acquiring a new language. It is also consistent with the limited amount of research (Tomoschuk et al., 2021) suggesting that this non-native language interference might persist even after the initial stages of L3 acquisition. Here we show for the first time that increased L3 interference can disrupt L2 production in trilinguals who acquired all three languages early in life and who have an intermediate proficiency in both of their non-native languages. Furthermore, most of the acquisition literature has focused on interference stemming from a non-native language (L2) acquired before the other non-native language (L3). Tomoschuk et al. (2021) only observed increased non-native interference on an L3 but not on a (more dominant) L2. Here we show that this interference between non-native languages can also influence the non-native language that was acquired first (L2), potentially provided that this L2 has not reached very high levels of proficiency or use.

One potential mechanism leading to increased interference between non-native languages might be that trilinguals apply more inhibition (or apply inhibition more successfully or efficiently) over their native language. The interpretation that trilinguals suppress their L1 most strongly is in line with Green's (1998) inhibitory control hypothesis arguing that the amount of inhibition applied is relative to the proficiency in that language. As a consequence of the increased use of inhibition, interference from the native language might be reduced when using a less proficient language.

Indeed, our trilingual participants had easier access in the rhyme task to L3 translation equivalents than L1 equivalents after L2 use, suggesting that they suppressed the L1 more strongly than the L3. These language differences were not observed in the baseline task, suggesting that they were not due to differences between the L1 and L3 target words or between the probe-target relationships. They were not observed on within-language trials that were named in the rhyme language either. Multilinguals might apply language control not only by inhibiting a non-target language but also by over-activating the less proficient languages (L2/L3, Philipp, Gade, & Koch, 2007). If the observed across-language L1/L3 differences in the rhyme task were due to L3 words being activated more strongly than L1 words, language differences should occur on the within-language trials too. The finding that only across-language trials showed a language difference suggests that this language effect was related to reduced L1 accessibility rather than increased L3 accessibility.

Various other variables have been suggested to explain increased interference between non-native languages. The L2 and L3 might be more cognitively similar (e.g., Bardel & Falk, 2007) if they are acquired in similar circumstances. In Experiment 1, however, only the L3 was acquired as a true "classroom" language. The L2 was, on average, acquired during the first three years of life and is furthermore part of the bilingual society formed by the Basque Country. Furthermore, the suggestion that non-native language interference might be related to the non-native languages being taught in the L1 (Tomoschuk et al., 2021) does not hold here considering that most participants attended education in L2 only or in a combined L1/L2 system and used Basque across their educational programme rather than as a classroom language taught through another language. Typological proximity too is unlikely to explain the results considering that the L2 differed substantially from both the L1 and L3 but, if anything, shows more phonological and orthographic overlap with the L1.

Experiment 1 thus suggests that the interference between non-native languages might not be related to cognitive similarity between the L2 and L3 but rather to the way trilinguals apply more L1 than L3 inhibition. However, further analyses showed no correlation between the relative number of L1 intrusions and the relative accessibility of L1 words in the rhyme task. This could suggest that despite the same group of participants showing fewer L1 intrusions in the speeded naming task and reduced L1 access in the rhyme task, there is no direct relationship between the two findings. However, the two tasks are also very different in many other aspects (e.g., the speeded switching task requires naming of specific words while the rhyme task allows for free retrieval; the speeded switching task is a trilingual environment while the rhyme task was completed in single-language blocks). Furthermore, the rhyme task might elicit other individual differences beyond language production, including differences in the way people generate rhymes and the size of their vocabulary (i.e., the number of competitors for the target rhyme words). These many task-related differences could mask correlations within a relatively small sample. Therefore, while the rhyme task in Experiment 1 suggested that L1 translation equivalents were less accessible as a consequence of inhibition, we did want to investigate the potential role of inhibition and the relationship with intrusions further. In Experiment 2, we therefore firstly aimed to replicate the intrusion findings from the speeded naming task (in a different type of trilinguals) and we further examined the second question regarding L1 versus L3 inhibition. We used a different task (n-2 switching task, as explained below) to provide a more complete picture of the potential role of language inhibition.

Experiment 2

Introduction

N-2 switching task

Experiment 2 used the n-2 switching task to further examine the question of inhibition of the L1 versus L3. In this task, participants are asked to switch languages on every trial. The n-2 trial (i.e., two trials before the target) either has to be produced in the same language or in a different language than the current trial (e.g., L1-L2-L1 would be an n-2 repetition trial while L3-L2-L1 would be a switch trial). If the language used on trial n-2 has to be suppressed when switching to the L2 (trial n-1), participants should need more time to switch back to that language on trial n. In other words, if language X is suppressed when using the L2, it should take trilinguals more time to switch back to language X again (Lx-L2-Lx) than when a different language was used two trials ago (Lz-L2-Lx). Indeed, several trilingual switching studies using a range of language combinations have shown these n-2 costs, with trilinguals responding more slowly in sequences in which trial n and n-2 have to be named in the same language (repetition) than when they are named in different languages (e.g., Declerck & Philipp, 2018; Declerck, Thoma, Koch, & Philipp, 2015; Guo, Liu, Chen, & Li, 2013; Philipp et al., 2007).

These repetition costs are taken as a reflection of persisting inhibition of a previously used language. Importantly, if the L1 is suppressed more strongly than the L3 when using the L2, this n2-repetition cost should be larger for the L1 than for the L3: the difference between (L1-L2-L1) and (L3-L2-L1) should be larger than between (L3-L2-L3) and (L1-L2-L3). Some studies have indeed suggested that n-2 repetition costs are largest for the trilinguals' most dominant language(s) (e.g., Declerck et al., 2015; Philipp et al., 2007). Nevertheless, it should be noted that several studies have *not* shown larger repetition costs for the L1 than L2/L3 or for the L2 than L3 (e.g., Philipp & Koch, 2009; see Declerck & Koch, 2022, for a review showing inconsistencies across studies). Importantly, however, n-2 repetition effects (and potential language differences) are more likely to be explained by inhibition accounts than by over-activation of the target language. The latter would predict repetition priming rather than a cost (i.e., if the Lx is over-activated on trial n-2 and not suppressed at all, trilinguals should be faster on Lx-L2-Lx trials than on Lz-L2-Lx trials). Examining L1 versus L3 repetition costs thus allowed us to examine potential differences between L1 and L3 *suppression*. Importantly, we always kept the L2 as the middle trial (n-1) to specifically examine L1 versus L3 inhibition applied while switching to the L2 (contrary to previous studies, in which the middle trial for L1 repetition costs could be either L2 or L3 while the middle trial for L3 costs could be either L2 or L1). If more L1 inhibition is applied than L3 inhibition during L2 naming, we would expect L1 repetition costs to be larger than L3 repetition costs.

The use of an n-2 switching task allowed us to examine suppression of L1 versus L3 through a different type of task that, in both the language-switching as well as the task-switching literature, is frequently used as a measure of dominant language/task inhibition (e.g., Declerck & Philipp, 2018; Philipp et al., 2007). This was intended to complement the, less frequently used, rhyme task from Experiment 1. Crucially, however, we designed the n-2 task differently than previous studies by always using the L2 as the n-1 trial, which allowed us to specifically look at L1 versus L3 inhibition during L2 naming. Furthermore, we assessed if and how n-2 costs were related to intrusions in the speeded naming task to elucidate the relationship between non-native language interference and inhibition. The n-2 switching task is more similar to the speeded switching task than the rhyme task, allowing us to examine this correlation in the absence of large task differences (as was the case in Experiment 1).

Item-specific versus global-language control

Additionally, this n-2 task allowed us to assess global inhibition (of the language as a whole) versus item-specific inhibition. When applying inhibition, multilinguals can do this reactively by just suppressing the translation equivalent of the target word in the non-target languages (e.g., when naming "caballo" in Spanish, bilinguals might just suppress the equivalent "horse" in English but no other English words). Alternatively or additionally, multilinguals might suppress all lemmas in the non-target language (e.g., not just "horse", but also "apple", "dress", etcetera). While previous studies (e.g., Philipp et al., 2007) have shown that n-2 costs can be used to study inhibition during language production, it is less clear if this inhibition is applied at a global level or is also related to specific items that have to be named. Whole-language and item-specific inhibition are not mutually exclusive: multilinguals might apply inhibition globally but also increase inhibition of the translation equivalents in an item-specific manner (e.g., Declerck & Philipp, 2017). Item-specific effects have often been studied by repeating the same pictures in a picture-naming task. For example, Misra et al. (2012) showed a facilitation effect when the same pictures had to be named in the L2 after a block of L1 naming. However, this benefit of repeating the same pictures was not present in L1 after L2 naming, suggesting inhibition of L1 forms while naming the pictures in the L2. However, this

Table 2
Summary of the objective and subjective measures of language proficiency, language exposure, and language use for English, French, and Spanish.

	English			French			Spanish		
	Mean	SD	Range	Mean	SD	Range	Mean	SD	Range
Age of Acquisition	0	0	0–0	10.1	4.1	1–28	13.5	5.3	1–25
LexTALE[1]	X	X	X	69.2	12.4	45–98	69.7	13.4	43–100
Self-rated proficiency (0–10)									
Speaking	10.0	0.0	10–10	6.3	2.0	2–10	6.9	1.9	2–10
Understanding	9.9	0.2	9–10	7.6	1.6	3–10	7.9	1.5x	4–10
Writing	9.9	0.3	8–10	6.5	2.0	2–10	7.1	2.0	2–10
Reading	9.9	0.2	9–10	7.7	1.6	4–10	8.1	1.5	4–10
%exposure[2] (0–100)	89.0	16.4	35–100	24.6	27.6	0–100	27.3	27.5	0–100
%speaking[2] (0–100)	89.5	15.8	35–100	18.3	25.5	0–100	20.9	27.2	0–100

[1] LexTALE score is calculated as the percentage of words identified correctly (not including non-words).
[2] Participants were asked to rate exposure from “never” (0%) to “all the time” (100%), based on the past six months. While we asked participants to make sure their scores added up to 100% across the three languages, not all participants did this (i.e., some interpreted 100% as meaning speaking Lx a lot but not exclusively).

study did not include a direct comparison between item-specific and whole-language suppression.

This comparison was addressed by Branzi, Martin, Abutalebi, and Costa (2014), who asked highly proficient bilinguals to name pictures in three blocks, using the order L1-L2-L1 or L2-L1-L2. The blocks included new pictures as well as pictures that had also been named in the other language in the other block. The L1 was affected more negatively by previous L2 naming than the L2 naming was by previous L1 naming, again suggesting that bilinguals applied more control over the L1 while using the L2 than vice versa. While the exact pattern of results differed, both new and repeated items showed evidence for more L1 than L2 control, suggesting that this control was applied at the whole-language level. Similarly, Degani, Kreiner, Ataria, & Khateeb (2020) also found increased L1 error rates after L2 exposure for both repeated and new items, providing evidence for whole-language control. However, repeated items did show more cross-language intrusions than new items, suggesting item-specific control might go beyond the whole-language level. Furthermore, Van Assche, Duyck, & Gollan (2013) suggested that whole-language control might be applied differently depending on the type of bilinguals. In their study, participants were given a verbal fluency task and had to produce as many words as possible starting with a given phoneme, which was either the same across languages or different. Two participant groups (late L1-dominant Dutch-English bilinguals and early mostly L2-dominant Chinese-English bilinguals) both showed item-specific control effects in the dominant language, which elicited fewer correct response if the same category had just been used in the non-dominant language. Whole-language control effects (i.e., worse performance if a different category was used in the other language first), however, were only observed for Chinese-English bilinguals. The current literature thus suggests that both whole-level and item-specific control might be applied over the more dominant language. However, it leaves open the question whether this also applies to unbalanced trilinguals. Furthermore, it remains unclear if control might be applied more strongly over translation equivalents, in addition to more global language control, and, if so, if this differs between L1 and L3 control processes.

In our Experiment 1, the rhyme task showed that the translation equivalents of L2 words were less accessible in the L1 than L3. However, a disadvantage of the rhyme task was that it only allowed us to assess accessibility of translation equivalents (i.e., the words corresponding to the pictures participants had just named) but not the consequences for the lexicon more globally. The absence of language differences on the within-language trials could suggest that inhibition differences only affected translation equivalents. However, the within-language items had been used multiple times and only in that specific language in the naming task, which could lead to repetition benefits in both languages outweighing any whole-language suppression effects. This leaves open the question whether bilinguals inhibit specific translation equivalents or the L1 lexicon more globally while naming in L2. In Experiment 2, we therefore used an n-2 switching task to compare item-specific inhibition and more global language inhibition (i.e., inhibition of other items than the one currently used). We manipulated item repetition on trial n relative to n-1 (e.g., on n-1 participants had to name a picture of a horse in their L2 and on trial n they named the same picture in their L1 or L3 or they named different pictures on trials n-1 and n). On trial n-1 (French), trilinguals are argued to apply inhibition, in particular over the language used on trial n-2 (e.g., English). On the French n-1 trial, they might inhibit all words in English (n-2 language) equally strongly, or they might apply only/more inhibition over the English translation equivalent of the n-1 item in particular. To give an example, when participants have to say “horse” in French on n-1, they might inhibit all English words equally or they might inhibit “horse” most strongly. If trilinguals apply more global language inhibition (i.e., if they suppress multiple words in the other languages and not just the translation equivalent) while naming in L2, n-2 costs should occur regardless of item repetition (i.e., returning to English on trial n should be costly regardless of the item). If trilinguals only apply item-specific inhibition, n-2 costs should only occur when items are repeated (i.e., returning to English on trial n should be especially difficult when “horse” has to be named again as it was suppressed strongly on trial n-1 when naming that picture in French). If both global and item-specific control are applied but item-specific control is applied most strongly, n-2 costs should be found for repeated and unrepeated items but should be largest for repeated items (i.e., there are n-2 repetition costs for all items, but most strongly for the repeated “horse”). Crucially, this is specifically related to the n-2 language as trilinguals are argued to suppress the language used during trial n-2 most strongly during trial n-1. If trilinguals inhibit the L1 more strongly than the L3 globally, larger L1 than L3 n-2 costs should be observed across both repeated and unrepeated items. However, if trilinguals especially inhibit the L1 translation equivalent of the word that needs to be named in the L2 during trial n-1, L1 n-2 costs should be larger than L3 costs (or only present) when the item is repeated.

Aims Experiment 2

Experiment 2 thus had two aims. First, we aimed to replicate the findings observed in the speeded switching task in Experiment 1, which showed more L3 than L1 intrusions during L2 naming. We used a different type of trilinguals, in which the L1 and L3 were reversed (i.e., English was the participants’ L1 and Spanish the L3 in Experiment 2). This way, we also reversed any potential influences of differences in word length as a consequence of Spanish words being longer than English words. Second, we aimed to assess L1 versus L3 inhibition during L2 naming through the n-2 task. We examined whether (in line with Experiment 1) trilinguals applied more inhibition over the L1 than L3 during L2 naming, in which case we should observe larger L1 than L3 n-2 costs. In addition, we examined whether language control is applied across all words in a language or only or most strongly for translation equivalents (in which case, n-2 costs should be strongest when items are repeated). We also examined whether these types of control (global-language or item-specific) differ for the L1 and L3.

Methods

Participants

The final dataset included fifty English-French-Spanish trilinguals (10 male, *M*age = 29.7; *SD*age = 7.0, see "procedure" for the process of selecting and excluding participants). Similar to Experiment 1, we recruited participants who acquired their L1 from birth and had an intermediate proficiency in their other two languages. This sample size was based on power simulations (using simr) using an effect size two-third the size (.24) of the fixed effect (.37) observed in the speeded naming task in Experiment 1. The smaller effect size was chosen to allow for potentially noisier data given that the replication was conducted online. Simulations showed that twenty participants yielded over 80 % power to replicate the speeded naming task results with this adjusted effect size. Power analyses for the n-2 task were more difficult considering that we had no good indication of the effect size associated with L1 and L3 differences and potential interactions with item repetition. We therefore opted for a larger sample size (50), again also considering recruitment constraints associated with this participant profile. All apart from six participants were right-handed. All participants had normal or corrected-to-normal vision and no known neurological, reading, or hearing impairments. The study was approved by the Ethics Committee in the Department of Psychology at the University of York and complied with the guidelines of the Helsinki Declaration.

Participants' proficiency and language use were assessed through a questionnaire and the LexTALE (French: Brysbaert, 2013; Spanish: Izura, Cuetos, & Brysbaert, 2014). All participants acquired English as their first language and French and Spanish later in life (with a range of age of acquisitions, see Table 2). They had an intermediate proficiency level in Spanish and French. All participants were born in the UK or Ireland and most participants were living there at the moment of testing (five participants were currently living in a French- or Spanish-speaking country). The language background thus differed from Experiment 1, with most participants in Experiment 2 living in a monolingual society in which their first language (English) was dominant and French and Spanish were used less frequently. Thirty-four participants had spent at least three months living in a French-speaking country and twenty-nine in a Spanish-speaking country. There was furthermore a range of proficiency and language use, with French being the second language for some participants and the third for others (for consistency, and in line with the mean age of acquisition, we will refer to French as the L2 and to Spanish as the L3).

Design

The speeded switching task was similar to the one used in Experiment 1 and compared the number of L1 versus L3 intrusions during L2 non-switch trials.

The n-2 switching task used reaction times (RTs) as the DV, with Language of naming (L1/L3), Trial type (trial n-2 using the same language as trial n or not), and Item repetition (picture same as previous trial or not) as the within-subject IVs.

Materials

We selected 24 pictures (eight of which were used in the speeded naming task) that represented easy-to-name items with high-frequency words (see Appendix B). All were non-cognates. English and Spanish words were matched on frequency and number of letters (similar to Experiment 1, Spanish words had more phonemes and syllables than English words, but now they were the participants' L3 and L1 respectively). English-L1 and Spanish-L3 words were also matched on their Levenshtein distance to French-L2 words. For the words used in the speeded naming task, we ensured they started with different phonemes to allow us to score intrusions.

We adjusted the choice of language cues due to colour similarities between the French and English flags (which could have influenced cross-language intrusions). We therefore used a cue that combined half of the country flag with a country symbol (Tower Bridge, Eiffel Tower, and the Sagrada Familia). For each language, the country symbol was placed in a different position relative to the flag (on the left, in the middle, or on the right), to make sure the cues could easily be distinguished.

Procedure

Participants completed three sessions. Given that there was no anticipated effect of time between the sessions in this study (and considering that participants could take part online whenever convenient for them), the interval between sessions varied from a few days to a few months. Participants were invited through Prolific.co and completed all sessions on Gorilla.sc (Anwyl-Irvine, Massonié, Flitton, Kirkham, & Evershed, 2020). The first session was a pre-screening to ensure that participants were native speakers of English and could speak both French and Spanish (considering that the pre-screening questions on Prolific.co only allowed us to select people who spoke French and/or Spanish but not necessarily both). In the pre-screening, participants were asked about their age of acquisition and their overall self-rated proficiency (0–10) of French and Spanish. They also completed the LexTALE and a short picture-naming task that included the 24 items used in the experiment. Participants were only invited for the experiment if A) they indicated they could speak both French and Spanish, and B) they named at least 18/24 pictures correctly in both French and Spanish. We also used the pre-screening to exclude participants with audio issues (i.e., participants with recordings that were far too short or too noisy were not invited for the experiment). Of the 117 participants who completed the pre-screening, 77 were invited to take part in the study. Of those, 72 completed the first session and 50 successfully completed both sessions and were included in the analysis (5 participants did not respond to the invitation for the second part; for 16 participants the responses were not recorded well enough for their data to be processed; and one participant was excluded because of low accuracy in the second experiment, which did not leave any trials in one of the conditions).

In session 2, participants completed the speeded switching task. This was similar to the task described in Experiment 1. However, we reduced the number of trials to avoid the task being too long for an online study. Participants completed 584 trials (576 experimental trials plus eight trials that were excluded for being the first of the block after breaks): 128 L1 and 128 L3 trials (32 nonswitch and 96 switch) and 320 L2 trials (192 nonswitch and 128 switch). Each trial also lasted 300 ms longer than in Experiment 1 to allow for short pauses between recordings (to minimise issues with recordings not being saved well due to the pace of the task). Participants were shown the image and language cue for 1000 ms and then saw a fixation cross for 1200 ms. Two seconds were recorded (1 s while they were asked to name the picture and the first second of the fixation cross) and responses given within those two seconds were scored, even though participants were told they had to respond during the image display. This session lasted approximately 20–25 min.

In session 3, participants completed the (slower pace) N-2 switching task. Similar to the first task, participants were told to name the picture in response to the language cue. Twenty-four pictures were used and all trials were language switches. Participants completed 436 trials in total, of which 192 were the experimental L1 and L3 trials of interest. 96 trials were completed in L1 and 96 in L3, with half of them being n-2 repetitions (i.e., L1-L2-L1 and L3-L2-L3) and the other half being n-2 switches (i.e., L3-L2-L1 and L1-L2-L3). Within each condition, half of the

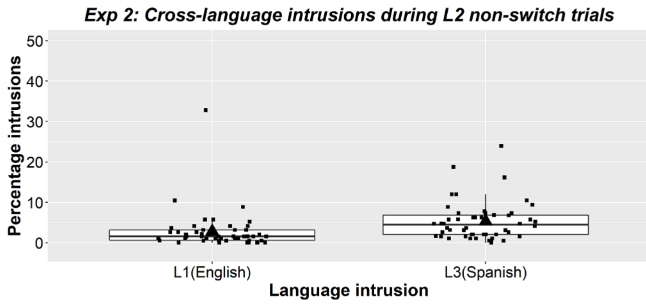

Fig. 5. Boxplots showing the percentage of language intrusions (L1 versus L3) during L2 non-switch trials. Each black square shows an individual participant (jittered). The horizontal line shows the median while the black triangle shows the mean.

Table 3
Means (and SDs) by language (English L1; Spanish L3), Trial type (Language n-2 repetition or switch), and Item type (repeated or different item).

	English (L1)		Spanish (L3)	
	Repeated item	Different item	Repeated item	Different item
Language n-2 repetition	1055.4 (145.7)	1174.1 (132.7)	1102.8 (142.6)	1157.0 (129.0)
Language n-2 switch	1042.3 (141.5)	1143.3 (147.6)	1081.8 (137.3)	1164.4 (113.3)

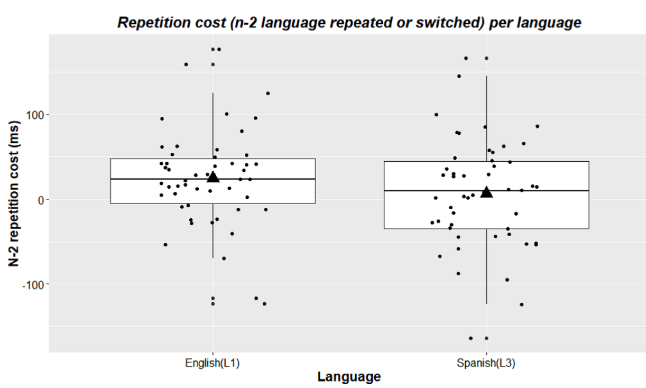

Fig. 6. Boxplots showing the N-2 repetition costs for the L1 (English) and L3 (Spanish). Each black dot shows an individual participant. The horizontal line shows the median while the black triangle shows the mean.

trials showed item repetitions (e.g., current and previous trial showing pictures of a duck). Each picture occurred once in each condition. The other trials were L2 trials (198) or L1/L3 filler trials (46) that included switches between L1 and L3 to avoid participants predicting the trial sequence. Prior to starting the task, participants completed a familiarisation phase in which they saw each picture with the corresponding words in each of the three languages (presented one at a time). They were asked to look at the picture and read the corresponding words. Participants also completed eight practice trials. The picture stayed on the screen for two seconds, followed by a fixation cross for 500 ms.

At the end of the third session, participants completed a questionnaire about their language use and background (see Table 2). The session lasted approximately 30 min.

Data analysis

The speeded naming task was scored and analysed in the same way as in Experiment 1. For the n-2 switching task, reaction times were scored using CheckVocal (Protopapas, 2007) and were log transformed to improve normality of the distribution (but untransformed means are presented throughout the Results section). A linear mixed-effect analysis was conducted with log RTs as the DV and the following fixed effects and their interactions: Naming Language (Spanish-L3: −0.5; English-L1: 0.5), Trial Type (n-2 repetition: −0.5; n-2 switch: 0.5), and Item Repetition (same item: −0.5; different item: 0.5). The model converged with intercepts for participants and items and by-participant slopes for trial type, language, item type, trial type × language, and language × item type.

Results

Speeded trilingual naming task

Across the 576 experimental trials[1] in the speeded trilingual naming task, participants on average answered 87.3 % correctly (SD = 10.0, range = 44–98 % correct). Participants did not respond on 2.5 % of trials (SD = 3.0). Cross-language intrusions were made on 9.5 % of trials (SD = 8.9) and within-language intrusions on 0.8 % of trials (SD = 0.9); 0.09 % of trials (SD = 0.3) were both cross- and within-language intrusions.

The error type of interest was the number of cross-language intrusions. We focused on the number of cross-language intrusions during French (L2) non-switch trials (see Fig. 5). During these trials, participants produced significantly more L3 (M = 5.3 %, SD = 4.8 %) than L1 (M = 2.8 %, SD = 4.8; β = 0.641, SE = 0.075, z = 8.590, $p < 0.001$) intrusions.

Similar findings were observed for French (L2) switch trials. Participants produced more L3 (M = 4.8 %, SD = 4.3) than L1 (M = 3.5 %, SD = 5.4) intrusions (β = 0.373, SE = 0.089, z = 4.198, $p < 0.001$). There was a significant effect of the language used on the previous trial (β = -0.259, SE = 0.089, z = -2.908, p = 0.004), which interacted with the language of intrusion (β = 0.816, SE = 0.178, z = 4.585, $p < 0.001$). Participants made more L3 (M = 5.2 %, SD = 4.9) than L1 (M = 2.4 %, SD = 4.1) intrusions when an L2 trial was preceded by an L3 trial but not when it was preceded by an L1 trial (L3 M = 4.5 %, SD = 4.4; L1 M = 4.6 %, SD = 7.2).

Given the lower number of trials to be named in L3 or in L1, and given our focus on L2 production, we did not further analyse the number of intrusions during L1 or L3 trials. Similar to Experiment 1, though, there were relatively more L3 intrusions during L1 trials (2.2 %, SD = 2.5) than L1 intrusions during L3 trials (1.6 %, SD = 4.2). Given that the task required participants to use the L2 more than the other languages, the largest number of intrusions came from the L2 (during L1 trials M = 10.0 %, SD = 8.0; during L3 trials M = 8.7 %, SD = 7.4).

In summary, and replicating Experiment 1, the speeded naming task shows that trilinguals were more likely to make L3 than L1 intrusions when having to name pictures in L2.[2]

N-2 switching task

In the n-2 switching task, we only analysed experimental trials that were named in English-L1 or in Spanish-L3. These trials were always preceded by an L2 trial and by either the L1 or L3 on trial n-2. Mean accuracy in L1 English was 92.5 % (SD = 9.1) and in L3 Spanish 83.7 % (SD = 13.5; some recordings were missing for two participants, mean accuracy is relative to number of recordings). We did not analyse

[1] For a few participants, a few trials were not recorded correctly. When computing percentages per participants, this was relative to the number of recorded trials for that participant, excluding trials with recording failures.

[2] In both Experiments, we assessed if the number of intrusions from each language changed during the task. Comparing the first and second half of the speeded naming tasks showed a slight increase (which was not significant) in overall number of intrusions in the second half in Experiment 1 but no difference across task halves in Experiment 2. Crucially, there was no interaction between task half and language of intrusions, suggesting that the difference between L1 and L3 intrusions was stable across the task and did not increase (or decrease) with time (see Supplementary Materials).

accuracy further given that trial type (language repetition or not) also depended on accuracy on the previous two trials. For RT analyses we therefore not only removed incorrect trials but also trials that were preceded by one or two incorrect responses (22.0 % of correct responses). RT outliers (2.5SD above/below the mean per participant and condition) were removed too (0.9 % of trials).

The RT means by condition are reported in Table 3 and the repetition costs are shown in Fig. 6. There was a significant effect of trial type, with slower responses when the language of trial n was the same as that of trial n-2 ($Mcost = 15.4$, $SD = 38.5$; $\beta = -0.014$, $SE = 0.005$, $t = -2.762$, $p = 0.008$). There was no main effect of language ($\beta = -0.017$, $SE = 0.012$, $t = -1.461$, $p = 0.150$) but trial type interacted with language ($\beta = -0.019$, $SE = 0.009$, $t = -2.029$, $p = 0.049$). There was a significant repetition cost in the L1, with slower responses if trial n-2 had been named in the L1 too ($Mcost$ 24.3, $SD = 56.4$; $\beta = -0.024$, $SE = 0.007$, $t = -3.480$, $p = 0.001$). However, there was no significant cost in L3 ($Mcost$ 6.6, $SD = 63.6$; $\beta = -0.006$, $SE = 0.007$, $t = -0.906$, $p = 0.370$, see Fig. 6). This suggests that while L1 words were named more slowly when the L1 had been used before switching to the L2, this was not the case for L3 words.

We also examined whether these findings were specific to the items just used. There was a main effect of item repetition ($\beta = 0.087$, $SE = 0.007$, $t = 12.076$, $p < 0.001$), with faster responses when the item was repeated on two consecutive trials ($M = 1068.1$, $SD = 126.3$) than when it was different from the previous trial ($M = 1160.9$, $SD = 115.5$). The effect of item repetition was present in both languages ($ps < .001$) but was larger for L1 ($Mdifference = 109.1$, $SD = 55.7$) than L3 trials ($Mdifference = 72.4$, $SD = 88.5$; $\beta = 0.031$, $SE = 0.012$, $t = 2.484$, $p = 0.017$, see Table 3). However, this item repetition effect did not interact significantly with trial type ($\beta = 0.001$, $SE = 0.009$, $t = 0.074$, $p = 0.941$) or with trial type and language ($\beta = -0.032$, $SE = 0.017$, $t = -1.806$, $p = 0.071$). This suggests that while item repetition influenced overall naming times, it did not influence the n-2 cost.

Correlation between intrusions and n-2 repetition cost

We also assessed whether there was a correlation between the number of L1 intrusions in the speeded naming task and the L1 n-2 repetition cost in the n-2 switching task to assess if our measure of inhibition was related to the intrusions. We were specifically interested in assessing whether people who made relatively few L1 intrusions (as compared to L3 intrusions) also had relatively large L1 n-2 repetition costs (as compared to L3 repetition costs, i.e., a negative relationship between intrusions and repetition costs). To do this, we used the percentage difference between L1 and L3 intrusions and the RT difference between the L1 and L3 n-2 repetition cost. As shown in Supplementary Fig. 2, participants with relatively fewer L1 intrusions were also the ones with relatively higher L1 n-2 repetition costs ($r = -0.333$, $p = 0.018$), suggesting that those who inhibited the L1 more strongly during L2 naming (in the n-2 switching task) also showed fewer L1 intrusions during L2 naming (in the speeded naming task).

Discussion

Experiment 2 firstly aimed to replicate the cross-language intrusion findings from Experiment 1. The speeded switching task again showed more L3 than L1 intrusions during L2 naming. Interestingly, this replication was observed with a different type of trilinguals in which the L1 and L3 were reversed relative to Experiment 1. Furthermore, the replication in Experiment 2 was conducted online, suggesting that these paradigms can be used to elicit and study cross-language intrusions in studies in the lab as well as online.

In addition, we examined inhibition of the L1 versus L3 using an n-2 switching task. N-2 repetition costs were larger for the L1 than the L3, suggesting that participants inhibited the L1 more than the L3 when switching to the L2. These findings are in line with previous studies (e.g., Declerck et al., 2015) who also found that n-2 costs were largest for the more dominant language(s) (but see Declerck & Koch, 2022, for a review). In our study we specifically manipulated the n-1 trial to always be the L2, thus focusing on L1 versus L3 differences while naming in the L2. In line with the rhyme task used in Experiment 1, these findings suggest that trilinguals indeed suppress their L1 more strongly than their L3 when using the L2. In Experiment 2, this was furthermore directly associated with the number of intrusions made in the speeded task. Participants who showed relatively larger L1 repetition costs (suggesting more L1 inhibition) also showed fewer L1 intrusions (i.e., L1 translation equivalents were less accessible). This suggests that trilinguals might indeed experience more interference between non-native languages than from a native language because they suppress their native, more proficient, language more strongly.

As an additional and novel question, Experiment 2 manipulated item repetition to assess whether language control was applied globally across the language or on specific items (i.e., translation equivalents of the word that had to be used). Although item repetition led to faster naming times overall, it did not interact with the n-2 cost. This suggests that (L1) language control was applied globally across multiple items and not just at the level of specific items. This is in line with findings observed by Branzi and colleagues (2014). Van Assche et al. (2013) only observed whole-language inhibition for bilinguals with a high proficiency in both languages. Our findings suggest that these findings can also extend to trilinguals with a clear difference in dominance between their L1 and L2/L3. It should be noted here that both repeated and non-repeated items were used in all three languages within the same task. It is therefore possible that the participants applied control globally but only on the words that had the highest frequency in this context. Despite this experiment not allowing us to conclude that the *entire* lexicon was suppressed, it does show that trilinguals suppress words beyond translation equivalents of the previously used L2 word. Our findings furthermore suggest that, contrary to Degani et al. (2020), our trilingual participants did not apply additional item-specific control. Translation equivalents of just-named items as well as those of the other items were suppressed equally during L2 naming. The main effect of item repetition could suggest that *less* control was applied on translation equivalents. However, this interpretation is less likely given that item repetition only influenced overall RTs and did not interact with our measure of control (n-2 costs). Rather, the main effect of items might reflect repetition priming (i.e., facilitation because the same picture is presented again). Crucially, this item repetition priming occurred during both n-2 repetitions and n-2 switch trials.

General discussion

Across two experiments (Experiment 1: Spanish-Basque-English trilinguals; Experiment 2: English-French-Spanish trilinguals), we examined (control of) interference between two non-native languages. Using a speeded naming task, both experiments showed that trilinguals made more L3 than L1 intrusions during L2 naming. This suggests that while using a non-native language, trilinguals experienced more interference from another non-native language than from a native language. We also examined whether this interference was related to how strongly participants inhibited their L1 versus L3 during L2 naming. Using two different tasks, both experiments suggested that trilinguals suppressed their L1 more strongly than their L3. Furthermore, this applied to both translation equivalents of just-used pictures as well as more globally to other items in the lexicon. In Experiment 2, the relative amount of L1/L3 inhibition was related to the relative number of L1/L3 intrusions, suggesting that the interference trilinguals experience between non-native languages is related to weaker inhibition of a non-native than native language.

Language interference

Using two different language groups (with the L1 and L3 reversed), both experiments showed more L3 than L1 intrusions during L2 naming, despite the L1 being the more proficient and most used language by far. These findings are compatible with the L3-acquisition literature often showing an influence of the L2 on L3 acquisition (e.g., Mickan et al., 2020; Puig-Mayenco et al., 2020). Here we show that this close relationship between non-native languages extends to trilinguals far beyond the initial stages of language acquisition. In both Experiments, the vast majority of participants started L2 and L3 acquisition during childhood or early adolescence, in many cases more than a decade before study participation. They had also reached an intermediate proficiency level in both non-native languages. Even in these trilinguals, the non-native languages continued to influence each other more strongly than the native language. We furthermore show that this has noticeable consequences for non-native language production, which was negatively affected by cross-linguistic intrusions from the other non-native language. We focused on the consequences of non-native language intrusions for another non-native language, where the intrusions were indeed most pronounced. This is likely because the L2 was of a much lower proficiency level than the L1, potentially making this language more susceptible to intrusions than the L1. Remarkably, though, both experiments suggested that even the L1 can be affected by L3 intrusions, with the most proficient language (L1) experiencing more intrusions from the least-proficient language (L3) than vice versa.

The two experiments furthermore showed that this L3 interference can be present in different types of trilinguals. The participants in Experiment 1 mostly acquired all three languages during childhood. Their L1 and L2 were both present in the local community while the L3 remained mostly a classroom language. Their L2 differed typologically from the L1 and L3 but, if anything, was more similar to the L1. In contrast, most participants in Experiment 2 started L2 and L3 acquisition slightly later (at the start of adolescence). Their L2 and L3 were labelled as such based on mean Age of Acquisition but L2 and L3 mean proficiency and exposure were comparable to each other while participants in Experiment 1 were typically more proficient in and had more exposure to the L2 than L3 (see Tables 1 and 2). Almost all of the participants in Experiment 2 were living in an L1-dominant society that offers relatively low use of the L2 and L3 (although many participants had previously spent some time living in an L2- or L3-associated country). The L2 and L3 were more similar typologically to each other than to the L1. This type of trilingual participants is more similar to the type of participants studied in the L3-acquisition literature. Nevertheless, the findings in terms of cross-language intrusions and interference stemming from the L3 were very similar in the two Experiments. This consistency, despite the clear differences between populations, together with the many differences between the L2 and L3 in Experiment 1 also rule out various factors related to L2/L3 similarity that have been suggested to underly L2/L3 interference (e.g., Bardel & Falk, 2007).

L1 versus L3 control

Instead we show evidence that language interference and intrusions might be related to how strongly trilinguals control their L1 and L3. Studies with bilinguals suggest that inhibition is an important part of language control (Green, 1998) and is applied to avoid interference from the non-target language. This control might be applied more strongly over the L1 to allow for use of the L2 than vice versa (e.g., de Bruin et al., 2014; Levy et al., 2007; Meuter & Allport, 1999; Misra et al., 2012). Here we show that after L2 production, L1 words are less accessible than L3 words. In Experiment 1, L1 translation equivalents were less likely to be used as rhyme words than L3 equivalents after L2 naming. In Experiment 2, trilinguals showed larger L1 n-2 costs than L3 costs. Both findings suggest that trilinguals indeed apply more inhibition over their L1 than over their L3 and that this suppression is administered both at the level of specific items (translation equivalents, Experiment 1) as well as globally across other words in the non-target language (Experiment 2). Successful L1 inhibition (fewer L1 intrusions) might, however, depend on the preceding language. Both Experiments (although this interaction with language of the previous trial did not reach significance in Experiment 1) suggested that L1 interference was more likely immediately after L1 use (i.e., when switching from the L1 to the L2), possibly because preceding L1 use required lifting L1 inhibition. Furthermore, in Experiment 2, L1 control measured through the n-2 task was directly associated with the cross-language intrusions in the speeded task. While these correlations cannot establish a causal direction, it does suggest that trilinguals experience more interference from an L3 than from an L1 because they suppress their L3 less.

Crucially, both Experiments suggest that L1-L3 differences in interference are closely related to *inhibition* of a more proficient language. Alternative theories have proposed that multilinguals apply language control in the form of *(over-)activating* the less proficient language (e.g., Philipp et al., 2007). In Experiment 1 (rhyme task), such over-activation should have resulted in higher L3 performance across conditions rather than the observed L1-L3 differences in one specific condition (across-language trials). In Experiment 2 (n-2 switching task), over-activation should have led to n-2 repetition benefits instead of costs. Together, these two experiments highlight the role of inhibition during multilingual language control and suggest that inhibition, rather than over-activation, could explain L3 intrusions during L2 production.

These experiments are the first to show a direct association between non-native language control and interference (cross-language intrusions) in trilinguals but leave open questions about the exact ways in which language control differs between an L1 and L3. One explanation is that trilinguals always apply language control/inhibition relative to the proficiency of the language (Green, 1998). Following this explanation, trilinguals might “simply” suppress the L3 less strongly because it is the less proficient language, even if though this can have negative consequences in some contexts (in this case, leading to more intrusions in the speeded naming task). However, another (not mutually exclusive) explanation is that a trilingual’s control is more well-developed for an L1 based on the experience they have suppressing this language (in line with the Adaptive Control Hypothesis, Green & Abutalebi, 2013). Trilinguals might be used to suppressing their L1 whenever they have to or want to use an L2 or L3. Participants in Experiment 1 might need to suppress their L1 Spanish whenever they want to speak their L2 Basque with a Basque-dominant speaker or when they use their L3 English in the classroom. Participants in Experiment 2 might need to suppress their L1 English whenever they use their L2 or L3 with French- or Spanish-speaking friends or when they travel. However, trilinguals might not necessarily need much L3 control in their daily lives while they are using the L1, given that the L1 might be less susceptible to interference due to its high proficiency. As a consequence, the trilinguals tested in our study might have more experience controlling their L1 than L3 and might therefore be able to control their L1 more successfully. This latter explanation suggests that trilinguals who do have more “practice” controlling their L3 in their daily lives might be more successful at suppressing their L3 and avoiding intrusions when a context requires this. For example, a trilingual who has an intermediate proficiency in their L2 but who uses this language very frequently in an L2-dominant environment might also develop more efficient L3 control to avoid L3 intrusions during L2 use. Future research will need to examine if trilingual speakers of less-proficient non-native languages always apply less L3 control or if L3 control can be developed flexibly depending on the control needs required by the language environment.

Conclusion

While most research on language control has focused on a bilingual’s L1 and L2, here we show the importance of studying how trilinguals control their L3. Despite the higher proficiency in and use of their L1, trilingual L2 production experienced more interference from the L3 than L1. These cross-language intrusions can cause noticeable disruptions during L2 production. We furthermore show that this L3 interference is related to how well trilinguals control their L3 as compared to their L1. Interference between non-native languages thus continues beyond early stages of language acquisition into proficient trilingual speakers.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data and analysis script for both Experiments are available on https://osf.io/wmehd/. The stimuli are provided in the appendices.

Acknowledgements

This research was supported by the Basque Government through the BERC 2022-2025 program and by the Spanish State Research Agency through BCBL Severo Ochoa excellence accreditation CEX2020-001010-S. LJH received funding from the Spanish government Juan de la Cierva-Incorporación Postdoctoral Grant. CDM received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant Agreement No: 819093), the Spanish Ministry of Economy and Competitiveness (PID2020-113926 GB-I00), and the Basque Government (PIBA18_29).

Appendix A. . Stimuli Experiment 1

Experiment 1 included eight pictures in the speeded naming task (Table A1). Spanish and English words were matched on frequency per million (English $M = 62$, $SD = 49$; Spanish $M = 33$, $SD = 27$, $t(7) = 1.583$, $p = 0.158$), number of letters (English $M = 4.5$, $SD = 0.5$; Spanish $M = 4.3$, $SD = 1.2$, $t(7) = 0.683$, $p = 0.516$), and number of phonemes (English $M = 3.4$, $SD = 0.5$; Spanish $M = 4.1$, $SD = 0.8$, $t(7) = -2.049$, $p = 0.080$). Spanish words were slightly longer in terms of syllables (English $M = 1.1$, $SD = 0.4$; Spanish $M = 2.0$, $SD = 0.5$, $t(7) = -2.966$, $p = 0.021$). Spanish and English words were also matched in terms of Levenshtein Distance to the Basque equivalent, which was computed as the number of edits needed to change one word into the other, relative to the length of the word (with 1 meaning complete overlap and 0 meaning no overlap: English $M = 0.04$, $SD = 0.08$; Spanish $M = 0.2$, $SD = 0.1$, $t(7) = -2.257$, $p = 0.059$).

The items used in the rhyme task were matched in terms of frequency of the target for the comparisons English within-language versus Spanish within-language (i.e., rhyme probes corresponding to words named in the same language in the naming task); English across-language versus Spanish across-language (i.e., rhyme probes corresponding to the translation equivalents of the words named in Basque in the naming task); English-within versus English-across; and Spanish-within versus Spanish-across (Table A2). Target length (number of letters, phonemes, and syllables) was matched for English-within versus English-across words and for Spanish-within and Spanish-across words. Spanish words were longer than English words in both the across- and the within-condition.

The probe and the relationship between the probe and target were also matched across conditions in terms of probe frequency, the Levenshtein Distance between probe and target, the number of possible high-frequency responses to the rhyme probe (with high-frequency defined as > 5 per million), and the number of other responses to the rhyme probe with a higher frequency than the target (e.g., if there were two other responses to "spoon" that were higher in frequency than the target "moon", the number would be "2"). In all conditions, at least 18/20 target-probe pairs were similar orthographically (e.g., "witch-switch" overlap phonologically and orthographically).

We also ensured that the rhyme probe and target were of similar length (apart from one pair, all pairs had a length difference of two letters at most). We selected probe words that were not semantically related to the target word (with one exception in the across-English condition). All rhyme targets and probes are shown in Table A3.

Table A1
Stimuli used in the speeded naming task in Experiment 1.

Spanish (L1)	Basque (L2)	English (L3)
Luna	Ilargi	Moon
Miel	Ezti	Honey
Pato	Ahate	Duck
Vaso	Edalontzi	Glass
Muro	Horma	Wall
Caballo	Zaldi	Horse
Ala	Hegal	Wing
Nube	Laino	Cloud

Table A2
Characteristics (means and SDs) of the targets and probes used in the rhyme task in Experiment 1 for the English within-language, Spanish within-language, English across-language, and Spanish-across language conditions. We used the EsPal database for Spanish (Duchon et al., 2013) and the MRC Psycholinguistic Database for English (Wilson, 1988).

	English within-language	Spanish within-language	English across-language	Spanish across-language
Frequency target*	50.8 (41.2)	44.4 (55.5)	48.1 (35.3)	37.7 (45.2)
Number of target syllables**	1.1 (0.2)	2.3 (0.4)	1.1 (0.2)	2.2 (0.5)
Number of target phonemes**	3.2 (0.5)	4.9 (0.9)	3.2 (0.5)	4.6 (1.2)
Number of target letters**	4.4 (0.7)	5.0 (0.9)	4.3 (0.7)	4.9 (1.0)
Number of possible rhyme responses with a high frequency*	10.7 (5.4)	9.9 (5.6)	8.4 (6.1)	9.7 (6.5)
Number of rhyme alternatives with higher frequency than target*	4.2 (3.2)	3.3 (2.9)	2.7 (2.3)	4.5 (4.6)
Probe frequency*	58.2 (70.9)	31.4 (34.0)	59.6 (92.3)	42.8 (54.1)
Levenshtein Distance Probe-Target*	0.69 (0.10)	0.67 (0.12)	0.65 (0.11)	0.64 (0.12)

*All conditions were matched on these characteristics (all ps >.05).
**Target word length was matched for English within- and across-language conditions and for Spanish within- and across-language conditions (all ps >.05) but could not be matched across languages (ps <.05).

Table A3
Target stimuli and rhyme probes used in the rhyme task in Experiment 1 by condition (within- or across-language) and by language (English or Spanish).

Condition	Language	Target	Probe
Within	English	Lock	Clock
Within	English	Ice	Rice
Within	English	Shark	Spark
Within	English	Sword	Lord
Within	English	Fire	Hire
Within	English	Hill	Skill
Within	English	Nose	Pose
Within	English	Bird	Third
Within	English	Dress	Press
Within	English	Witch	Switch
Within	English	Heart	Smart

(*continued on next page*)

Table A3 (*continued*)

Condition	Language	Target	Probe
Within	English	Broom	Doom
Within	English	Thief	Chief
Within	English	Bag	Tag
Within	English	Bell	Cell
Within	English	Knife	Wife
Within	English	Star	Jar
Within	English	Farm	Charm
Within	English	Snake	Cake
Within	English	Wave	Cave
Across	English	Horse	Force
Across	English	Shirt	Dirt
Across	English	Moon	Spoon
Across	English	Honey	Money
Across	English	Wheel	Steel
Across	English	Fox	Ox
Across	English	Duck	Truck
Across	English	Glass	Mass
Across	English	Wall	Ball
Across	English	Neck	Wreck
Across	English	Wing	Sing
Across	English	Snow	Grow
Across	English	Egg	Peg
Across	English	Pig	Wig
Across	English	Beach	Teach
Across	English	Path	Lath
Across	English	Cloud	Proud
Across	English	Sheep	Jeep
Across	English	Horn	Torn
Across	English	Bone	Tone
Within	Spanish	Puerta	Huerta
Within	Spanish	Pico	Chico
Within	Spanish	Flecha	Brecha
Within	Spanish	Muñeca	Beca
Within	Spanish	Puente	Cuente
Within	Spanish	Conejo	Espejo
Within	Spanish	Toro	Foro
Within	Spanish	Araña	Cabaña
Within	Spanish	Barco	Marco
Within	Spanish	Nudo	Dudo
Within	Spanish	Trigo	Digo
Within	Spanish	Cama	Rama
Within	Spanish	Clavo	Pavo
Within	Spanish	Taza	Raza
Within	Spanish	Vaca	Saca
Within	Spanish	Silla	Capilla
Within	Spanish	Cadena	Arena
Within	Spanish	Boca	Loca
Within	Spanish	Cabeza	Empieza
Within	Spanish	Llave	Nave
Across	Spanish	Caballo	Fallo
Across	Spanish	Camisa	Sonrisa
Across	Spanish	Luna	Cuna
Across	Spanish	Miel	Piel
Across	Spanish	Rueda	Pueda
Across	Spanish	Zorro	Gorro
Across	Spanish	Pato	Trato
Across	Spanish	Vaso	Paso
Across	Spanish	Muro	Duro
Across	Spanish	Cuello	Sello
Across	Spanish	Ala	Gala
Across	Spanish	Nieve	Atreve
Across	Spanish	Huevo	Llevo
Across	Spanish	Cerdo	Izquierdo
Across	Spanish	Playa	Vaya
Across	Spanish	Camino	Molino
Across	Spanish	Nube	Tuve
Across	Spanish	Oveja	Oreja
Across	Spanish	Cuerno	Tierno
Across	Spanish	Hueso	Peso

Appendix B

Spanish and English words were matched on frequency per million (set of 8: English $M = 66$, $SD = 76$; Spanish $M = 45$, $SD = 47$, $t(7) = 1.592$, $p = 0.155$; set of 24: English $M = 50$, $SD = 63$; Spanish $M = 38$, $SD = 59$, $t(23) = 1.925$, $p = 0.067$) and number of letters (set of 8: English $M = 4.1$, $SD = 1.1$; Spanish $M = 5.1$, $SD = 1.0$, $t(7) = -2.160$, $p = 0.068$; set of 24: English $M = 5.0$, $SD = 1.6$; Spanish $M = 5.5$, $SD = 1.1$, $t(23) = -1.801$, $p = 0.085$). Spanish words were longer in terms of syllables (set of 8: English $M = 1.1$, $SD = 0.4$, Spanish $M = 2.1$, $SD = 0.4$, *t*-test could not be conducted as the variance of difference was 0; set of 24: English $M = 1.4$, $SD = 0.6$; Spanish $M = 2.4$, $SD = 0.6$, $t(23) = -8.177$, $p < 0.001$) and phonemes (set of 8: English $M = 3.1$, $SD = 0.4$; Spanish $M = 4.9$, $SD = 1.1$, $t(7) = -4.249$, $p = 0.004$; set of 24: English $M = 3.8$, $SD = 1.2$; Spanish $M = 5.3$, $SD = 1.2$, $t(23) = -5.146$, $p < 0.001$). Spanish and English words were also matched in terms of Levenshtein Distance to the French equivalent, which was computed as the number of edits needed to change one word into the other, relative to the length of the word (with 1 meaning complete overlap and 0 meaning no overlap: set of 8: English $M = 0.06$, $SD = 0.1$; Spanish $M = 0.06$, $SD = 0.1$, $t(7) = 0.026$, $p = 0.980$; set of 24: English $M = 0.1$, $SD = 0.1$; Spanish $M = 0.1$, $SD = 0.2$, $t(23) = 0.936$, $p = 0.359$).

(See Table B1).

Table B1
Stimuli used in Experiment 2. The first eight items (in italics) were used in the speeded naming task; all 24 items were used in the n-2 switching task.

English	French	Spanish
duck	*canard*	*pato*
dog	*chien*	*perro*
cheese	*fromage*	*queso*
leg	*jambe*	*pierna*
skirt	*jupe*	*falda*
bed	*lit*	*cama*
apple	*pomme*	*manzana*
meat	*viande*	*carne*
candle	bougie	vela
doll	poupée	muñeca
hanger	cintre	percha
shoulder	épaule	hombro
snail	escargot	caracol
leaf	feuille	hoja
frog	grenouille	rana
rabbit	lapin	conejo
coat	manteau	abrigo
watch	montre	reloj
sheep	mouton	oveja
bird	oiseau	pájaro
butterfly	papillon	mariposa
rain	pluie	lluvia
suitcase	valise	maleta
city	ville	ciudad

Appendix C. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jml.2022.104386.

References

Anwyl-Irvine, A. L., Massonié, J., Flitton, A., Kirkham, N. Z., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioural experiment builder. *Behavior Research Methods, 52*(1), 388–407.

Bardel, C., & Falk, Y. (2007). The role of the second language in third language acquisition: The case of Germanic syntax. *Second Language Research, 23*(4), 459–484.

Bardel, C., & Falk, Y. (2012). The L2 status factor and the declarative/procedural distinction. In J. Cabrelli Amaro, S. Flynn, & J. Rothman (Eds.), *Third language acquisition in adulthood* (pp. 61–78). Amsterdam: John Benjamins.

Bardel, C., & Sánchez, L. (2017). The L2 status factor hypothesis revisited: The role of metalinguistic knowledge, working memory, attention and noticing in third language learning. In T. Angelovska, & A. Hahn (Eds.), *L3 Syntactic Transfer: Models, new developments and implications* (pp. 85–101). Amsterdam: John Benjamins.

Bogulski, C. A., Bice, K., & Kroll, J. F. (2019). Bilingualism as a desirable difficulty: Advantages in word learning depend on regulation of the dominant language. *Bilingualism: Language and Cognition, 22*(5), 1052–1067.

Branzi, F. M., Martin, C. D., Abutalebi, J., & Costa, A. (2014). The after-effects of bilingual language production. *Neuropsychologia, 52*, 102–116.

Brysbaert, M. (2013). Lextale_FR a fast, free, and efficient test to measure language proficiency in French. *Psychologica Belgica, 53*(1), 23–37.

De Bruin, A., Roelofs, A., Dijkstra, T., & FitzPatrick, I. (2014). Domain-general inhibition areas of the brain are involved in language switching: FMRI evidence from trilingual speakers. *NeuroImage, 90*, 348–359.

De Bruin, A., Carreiras, M., & Duñabeitia, J. A. (2017). The BEST dataset of language proficiency. *Frontiers in Psychology, 8*, 522.

Declerck, M., Grainger, J., & Hartsuiker, R. J. (2021). Proactive language control during bilingual sentence production. *International Journal of Bilingualism, 25*(6), 1813–1824.

Declerck, M., & Koch, I. (2022). The concept of inhibition in bilingual control. *Psychological Review*.Declerck, M., & Philipp, A. M. (2017). Is there lemma-based language control? The influence of language practice and language-specific item practice on asymmetrical switch costs. *Language, Cognition and Neuroscience*, *32*(4), 488-493.

Declerck, M., & Philipp, A. M. (2018). Is inhibition implemented during bilingual production and comprehension? n-2 language repetition costs unchained. *Language, Cognition and Neuroscience, 33*(5), 608–617.

Declerck, M., Thoma, A. M., Koch, I., & Philipp, A. M. (2015). Highly proficient bilinguals implement inhibition: Evidence from n-2 language repetition costs. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(6), 1911–1916.

Degani, T., Kreiner, H., Ataria, H., & Khateeb, F. (2020). The impact of brief exposure to the second language on native language production: Global or item specific? *Applied Psycholinguistics, 41*(1), 153–183.

Duchon, A., Perea, M., Sebastián-Gallés, N., Martí, A., & Carreiras, M. (2013). EsPal: One-stop shopping for Spanish word properties. *Behavior Research Methods, 45*(4), 1246–1258.

Duñabeitia, J. A., Crepaldi, D., Meyer, A. S., New, B., Pliatsikas, C., Smolka, E., & Brysbaert, M. (2018). MultiPic: A standardized set of 750 drawings with norms for six European languages. *Quarterly Journal of Experimental Psychology, 71*(4), 808–816.

European Commission (2012), *Special Eurobarometer 386: Europeans and their Languages.* [online] Available at: https://europa.eu/eurobarometer/surveys/detail/1049.

Falk, Y., & Bardel, C. (2011). Object pronouns in German L3 syntax: Evidence for the L2 status factor. *Second Language Research, 27*(1), 59–82.

Gollan, T. H., Schotter, E. R., Gomez, J., Murillo, M., & Rayner, K. (2014). Multiple levels of bilingual language control: Evidence from language intrusions in reading aloud. *Psychological Science, 25*(2), 585–595.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition, 1*(2), 67–81.

Green, D. W., & Abutalebi, J. (2013). Language control in bilinguals: The adaptive control hypothesis. *Journal of Cognitive Psychology, 25*(5), 515–530.

Guo, T., Liu, F., Chen, B., & Li, S. (2013). Inhibition of non-target languages in multilingual word production: Evidence from Uighur–Chinese–English trilinguals. *Acta Psychologica, 143*(3), 277–283.

Hirosh, Z., & Degani, T. (2021). Novel Word Learning Among Bilinguals Can Be Better Through the (Dominant) First Language Than Through the Second Language. *Language Learning, 71*(4), 1044–1084.

Izura, C., Cuetos, F., & Brysbaert, M. (2014). Lextale-Esp: A test to rapidly and efficiently assess the Spanish vocabulary size. *Psicológica, 35*(1), 49–66.

Jackson, G. M., Swainson, R., Cunnington, R., & Jackson, S. R. (2001). ERP correlates of executive control during repeated language switching. *Bilingualism: Language and Cognition, 4*(2), 169–178.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods, 44*(2), 325–343.

Levy, B. J., McVeigh, N. D., Marful, A., & Anderson, M. C. (2007). Inhibiting your native language: The role of retrieval-induced forgetting during second-language acquisition. *Psychological Science, 18*(1), 29–34.

Meuter, R. F., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language, 40*(1), 25–40.

Mickan, A., McQueen, J. M., & Lemhöfer, K. (2020). Between-language competition as a driving force in foreign language attrition. *Cognition, 198*, Article 104218.

Misra, M., Guo, T., Bobb, S. C., & Kroll, J. F. (2012). When bilinguals choose a single word to speak: Electrophysiological evidence for inhibition of the native language. *Journal of Memory and Language, 67*(1), 224–237.

Philipp, A. M., Gade, M., & Koch, I. (2007). Inhibitory processes in language switching: Evidence from switching language-defined response sets. *European Journal of Cognitive Psychology, 19*(3), 395–416.

Philipp, A. M., & Koch, I. (2009). Inhibition in language switching: What is inhibited when switching between languages in naming tasks? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(5), 1187–1195.

Protopapas, A. (2007). Check Vocal: A program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behavior Research Methods, 39*(4), 859–862.

Puig-Mayenco, E., González Alonso, J., & Rothman, J. (2020). A systematic review of transfer studies in third language acquisition. *Second Language Research, 36*(1), 31–64.

Rothman, J., & Cabrelli Amaro, J. (2010). What variables condition syntactic transfer? A look at the L3 initial state. *Second Language Research, 26*(2), 189–218.

Runnqvist, E., & Costa, A. (2012). Is retrieval-induced forgetting behind the bilingual disadvantage in word production? *Bilingualism: Language and Cognition, 15*(2), 365–377.

Tomoschuk, B., Duyck, W., Hartsuiker, R. J., Ferreira, V. S., & Gollan, T. H. (2021). Language of instruction affects language interference in the third language. *Bilingualism: Language and Cognition, 24*(4), 707–718.

Van Assche, E., Duyck, W., & Gollan, T. H. (2013). Whole-language and item-specific control in bilingual language production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(6), 1781–1792.

Williams, S., & Hammarberg, B. (1998). Language switches in L3 production: Implications for a polyglot speaking model. *Applied Linguistics, 19*(3), 295–333.

Wilson, M. (1988). MRC psycholinguistic database: Machine-usable dictionary, version 2.00. *Behavior Research Methods, Instruments, & Computers, 20*(1), 6–10.

Zheng, X., Roelofs, A., & Lemhöfer, K. (2020). Language selection contributes to intrusion errors in speaking: Evidence from picture naming. *Bilingualism: Language and Cognition, 23*(4), 788–800.