Research Article

María Martínez-Adrián* and María Nieva-Marroquín

Exploring the pseudo-longitudinal development of specific morphosyntactic features and syntactic complexity in CLIL young learners

https://doi.org/10.1515/iral-2021-0137
Received July 14, 2021; accepted January 6, 2022; published online January 20, 2022

Abstract: CLIL studies on language specific areas such as morphosyntax are still quite limited, particularly those with young learners. Likewise, the pseudo-evolution of morphosyntactic aspects across age and proficiency deserves particular attention. This paper will fill these gaps by investigating production accuracy and syntactic complexity in two different age/proficiency groups of CLIL primary school learners (grade 4 and 6). In terms of accuracy, L1 Spanish transfer effects associated with the pro-drop parameter (i.e. subject omission and subject-verb inversion), and the third person singular *-s* morpheme omission will be explored. Syntactic complexity will be operationalised through the production of simple and complex clauses. The findings obtained align with previous research in that the accumulated hours of CLIL + EFL exposure by grade 6 seem to positively affect the development of complexity measures. However, the lack of progress in the case of the rest of the features examined (i.e. subject omission, inversion of the subject and the third person singular *-s* omission) calls for the incorporation of focus-on-form components in CLIL programmes.

Keywords: CLIL; morphosyntax; young learners

***Corresponding author: María Martínez-Adrián**, Department of English, German, and Translation and Interpretation Studies, Faculty of Arts, University of the Basque Country UPV/EHU-Campus de Álava, Vitoria-Gasteiz, Spain, E-mail: maria.martineza@ehu.eus. https://orcid.org/0000-0002-0324-0443
María Nieva-Marroquín, Faculty of Arts, University of the Basque Country UPV/EHU-Campus de Álava, Vitoria-Gasteiz, Spain, E-mail: mnieva001@ikasle.ehu.eus. https://orcid.org/0000-0002-1069-3918

1 Introduction

Since the 1990s Content and Language Integrated learning (CLIL) has been implemented in primary/secondary education across Europe, and more recently in other parts of the world, as a way to provide a more intense and meaningful exposure to a foreign language than mere English as a foreign language (EFL) classes. Nowadays, these programmes are more the norm rather than the exception (see Pérez Cañado 2012 for a review).

Research studies on CLIL have proliferated in the last 10 years, particularly those targeting secondary school learners (e.g. Lasagabaster 2008; Merino and Lasagabaster 2018; Ruiz de Zarobe 2008, 2010). In the case of primary school learners, very little research has been done (e.g. Agustín Llach 2016; Fernández-Pena and Gallardo-del-Puerto 2021; Jiménez Catalán and Ojeda Alba 2009; Vraciu 2020), perhaps because CLIL programmes are not so widespread in primary education, as some scholars have noted (e.g. Pladevall-Ballester and Vallbona 2016), but also because this population has not been the target of investigations on second language (L2) learning up to quite recently (Martínez-Adrián et al. 2021; García Mayo and Imaz Agirre 2019; García Mayo and Lázaro Ibarrola 2015; Pinter 2007; Shintani 2012), despite the attested differences between adults and children in their approach to language learning (Oliver and Azkarai 2017). For example, children are prone to implicit learning, whereas older learners show more developed analytical and explicit learning abilities (e.g. Muñoz 2015). Given the uniqueness in how children engage in the L2 learning process, the examination of young learners is crucial if we want to provide this population with conditions most conducive to learning. Likewise, we need to make the most of the in-class exposure offered to them in low input contexts (Pinter 2011).

Even if recent years have witnessed a growing body of research on these meaning-oriented programmes, studies have primarily tackled the effect of CLIL exposure on the acquisition of general proficiency (Martínez-Adrián and Gutiérrez-Mangado 2015a; Jiménez Catalán et al. 2006; Lasagabaster 2008; Merino and Lasagabaster 2018; Ruiz de Zarobe 2008, 2010). However, research on language specific areas such as morphosyntax is still quite limited (Martínez-Adrián and Gutiérrez-Mangado 2009, 2015a, 2015b; Gutiérrez-Mangado and Martínez-Adrián 2018; García Mayo and Villarreal Olaizola 2010) and the vast majority have targeted secondary school learners, except for Fernández-Pena and Gallardo-del-Puerto (2021) and Vraciu (2020). Although Fernández-Pena and Gallardo-del-Puerto (2021) and Vraciu (2020) have investigated morphosyntactic features such as subject omission and verbal morphology, we are in the need of widening the scope of other features that have been the target of previous investigations conducted with CLIL secondary school learners.

Apart from the acute need to examine other morphosyntactic features in primary school learners in CLIL, the pseudo-evolution of morphosyntactic features across different age groups deserves particular attention. This is because the aforementioned investigations have primarily compared CLIL to NON-CLIL learners. This will lead us to the identification of problematic areas of language and to think about the most convenient measures that could promote a better development of linguistic features. By conducting pseudo-longitudinal studies, we will be contributing to the call made by Oliver and Azkarai (2017) for investigations in which participants with a range of ages are included and also to the call made in the CLIL field for the long term effect of these educational programmes (Pérez Cañado 2018).

Thus, the present study will address the aforementioned gaps by investigating a wider range of features, which will add to the contribution made by recent investigations (i.e. Fernández-Pena and Gallardo-del-Puerto 2021; Vraciu 2020). In particular, production accuracy as well as syntactic complexity measures will be examined following the trend of prior investigations conducted with CLIL secondary school learners (see Martínez-Adrián and Gutiérrez-Mangado 2015b). In terms of accuracy, L1 Spanish transfer effects associated with the pro-drop parameter (i.e. subject omission and subject-verb inversion)[1] (see 1 and 2 below), and the third person singular *-s* morpheme omission (see 3 below) will be explored. These linguistic targets have been found to pose difficulties for L1 Spanish learners of L2 English (i.e. Martínez-Adrián and Gutiérrez-Mangado 2009, 2015a; García Mayo 2003). Syntactic complexity will be operationalised through the production of simple and complex clauses, as in previous investigations with CLIL secondary school learners (i.e. Martínez-Adrián and Gutiérrez-Mangado 2015b; Lázaro Ibarrola and García Mayo 2012).

(1) Find the frog with a family (From Martínez-Adrián and Gutiérrez-Mangado 2009)

(2) Slept the baby for three hours (From García Mayo 2003)

(3) In the second the frog eh go out to the to the *tarro* ["jar"] (From Martínez-Adrián and Gutiérrez-Mangado 2015a)

More specifically, the present study will explore the pseudo-longitudinal evolution of these properties across two age groups of Spanish CLIL primary school learners: CLIL I ($n = 14$), from the 4th year of primary education; and CLIL II ($n = 12$), from the 6th year. This paper is organised as follows. Section 2 provides an overview of empirical findings related to general proficiency and morphosyntactic features in

1 Note that *that*-trace effects have also been examined in other investigations dealing with L1 transfer effects associated with the pro-drop parameter (García Mayo 2003).

CLIL programmes. Section 3 describes the methodology of the study. Results are offered and discussed in Sections 4 and 5, respectively. The paper finishes with the main conclusions drawn from the study.

2 Linguistic outcomes in CLIL programmes

The increasing interest in the CLIL approach and its implementation in education has resulted in a vast amount of research in many different countries. To mention but a few, Dalton-Puffer and Smit (2013), Dalton-Puffer (2011) and Marsh and Masih (1996) have offered comprehensive accounts of CLIL programmes in Europe. In Canada and the United States of America, the work by Celce-Murcia (1991) should not be overlooked. However, in order to limit the scope of this paper, it is necessary to outline the main research that has been conducted in Spain on the effect of CLIL on general proficiency and specific morphosyntactic features. The available evidence of the effect of CLIL exposure to the present date mainly comes from secondary school, with an ample majority of studies examining overall proficiency in English. Studies analysing specific language features are more limited.

2.1 General proficiency

Studies exploring general language proficiency in secondary education have confirmed that CLIL learners tend to outstrip same grade NON-CLIL learners and tend to perform as well as older NON-CLIL learners (Martínez-Adrián and Gutiérrez-Mangado 2015a; Lasagabaster 2008; Navés and Victori 2010; Ruiz de Zarobe 2010). These studies have examined the results in language proficiency tests comprising sections such as reading, listening, writing, use of English, vocabulary and speaking in secondary education In addition to these cross-sectional studies, other studies have tested the influence of CLIL on general proficiency from a longitudinal perspective (Merino and Lasagabaster 2018; Ruiz de Zarobe 2008, 2010), also revealing that the CLIL groups outperform NON-CLIL learners.

In primary school, the vast majority of the studies on general proficiency conducted have compared CLIL to NON-CLIL learners. In the study by Jiménez Catalán et al. (2006) carried out with 130 students from La Rioja and the Basque Autonomous Community, CLIL learners outperformed NON-CLIL learners when taking a cloze test designed to measure lexical, grammatical and discursive competence, as well as when taking a reading comprehension task and a receptive vocabulary level test. However, an advantage for CLIL learners was not attested in productive vocabulary measured by means of a composition task, as CLIL

learners produced fewer tokens and types than the NON-CLIL group. Productive vocabulary was not positively affected by CLIL in Jiménez Catalán and Ojeda Alba (2009) either. When performing a lexical availability task, NON-CLIL learners significantly produced more words than CLIL learners did. Lexical knowledge has also been examined in more recent studies. Agustín Llach (2016) compared a CLIL and a NON-CLIL group in their 4th, 5th and 6th grade, respectively. CLIL and NON-CLIL learners were receiving 105–110 h of EFL instruction on a yearly basis. In the case of CLIL learners, content lesson hours amounted to 72–74 every year. Learners were tested on lexical transfer and word frequency. In the case of lexical transfer, CLIL learners were found to be slightly better foreign language vocabulary users than NON-CLIL. In this regard, CLIL learners' lexical transfer behaviours were typical of more proficient learners, namely, fewer borrowings and more lexical creations. The examination of the results did not reveal a growing advantage of the CLIL approach over time, though.

Other studies have tackled receptive and productive skills jointly. Nieto Moreno de Diezmas (2016) tested 4th year CLIL and NON-CLIL learners in terms of writing, oral production and interaction, reading and listening comprehension. Both groups had studied English in infant education (270 h) and in primary education (450 h). The CLIL group had received 250 h of content lessons in English since the 1st year of primary education. No differences were observed between CLIL and NON-CLIL learners except for oral production and interaction in favour of CLIL learners.

This limited effect of CLIL in primary education is even more noticeable when studies control for in-class hours of exposure (one of the limitations of previous CLIL studies). When comparing 6th year CLIL to NON-CLIL learners matched in terms of hours of exposure (NON-CLIL = 210 h; CLIL = 156 h of EFL instruction + 54 CLIL hours), Pladevall-Ballester and Vallbona (2016) provided evidence of a better performance of the NON-CLIL group in listening and the inexistence of differences in reading comprehension. In this regard, the authors suggest that CLIL benefits might only be noticed in the longer term. As claimed by these authors, the initial cognitive challenge that young learners have to face when first exposed to CLIL directs their attention to meaning rather than to formal aspects of language. The results obtained in Pladevall-Ballester and Vallbona (2016) mirror the ones reported in Bret Blasco (2014) for productive skills where no differences emerged when comparing 5th year CLIL to NON-CLIL learners with a similar amount of in-class exposure. In this regard, these results are in line with the findings reported in other studies conducted in other European countries (Mattheoudakis et al. 2014; Serra 2007).

In order to contribute to the study of longitudinal effects of CLIL on foreign language outcomes, Pérez Cañado (2018) compared CLIL to NON-CLIL learners at the end of primary education (6th year) and four years later, when they were about to complete the last grade of compulsory secondary education (4th year). The results

obtained from the English language tests comprising the use of English, vocabulary, reading and speaking revealed the benefits of CLIL in all the categories examined in primary school learners. These differences between CLIL and NON-CLIL learners were even more remarkable in secondary school learners, giving support to the claim that the impact of CLIL seems to be greater in the long term.

2.2 Morphosyntactic features

A bulk of studies examining morphosyntactic features have focused on production accuracy and another strand on syntactic complexity, either alone or together with accuracy measures. The vast majority of investigations have been conducted with secondary school learners (Martínez-Adrián and Gutiérrez-Mangado 2009, 2015a, 2015b; Gutiérrez-Mangado and Martínez-Adrián 2018; García Mayo and Villareal Olaizola 2010; Lázaro Ibarrola 2012; Lázaro Ibarrola and García Mayo 2012), except for Fernández-Pena and Gallardo-del-Puerto (2021) and Vraciu (2020). Most investigations have compared CLIL to NON-CLIL learners and others, lower in number, offer a longitudinal perspective.

In the case of secondary education, those studies comparing CLIL to NON-CLIL learners exploring production accuracy in oral tasks have devoted their attention to L1 transfer effects, agreement morphology and the acquisition of articles in L2 English. Martínez-Adrián and Gutiérrez-Mangado (2009) examined data gathered by means of an oral storytelling task from 3rd year Basque/Spanish bilingual learners of English as a third language (L3) in CLIL and NON-CLIL contexts in order to test whether first language (L1) transfer effects on the use of null subjects, null objects, insertion of placeholders and negation would be minimised by participation in a CLIL programme. The results obtained at the time of data collection when CLIL and NON-CLIL learners had been exposed to 1,155 and 792 h of exposure, respectively, showed that CLIL learners significantly outperformed NON-CLIL learners only in the use of placeholders.

As far as the development of suppletive and affixal tense and agreement morphemes is concerned (i.e. third person singular *-s*, past tense *-ed* and auxiliary and copula *be*), García Mayo and Villarreal Olaizola (2010) found no significant differences between CLIL and NON-CLIL groups in their 3rd year of compulsory secondary education (CLIL = 875–910 h of exposure; NON-CLIL = 693 h of exposure) and 2nd year of Baccalaureate (CLIL = 1,443 h of exposure; NON-CLIL = 990 h of exposure) when examining a storytelling task. More recently, Martínez-Adrián and Gutiérrez-Mangado (2015a) and Martínez-Adrián and Gutiérrez-Mangado (2018) compared a 4th year CLIL group to a matching NON-CLIL group while keeping constant the variable of age at testing and the number of hours of exposure

to the target language (CLIL: 1,155 h of exposure; NON-CLIL: 1,148 h of exposure). In this regard, they could control for a similar amount of in-class exposure, which was lacking in previous research on the acquisition of morphosyntactic features by CLIL learners. They also compared the CLIL group to an older NON-CLIL group in the 2nd year of Baccalaureate with a slightly inferior amount of exposure (990 h). In particular, they examined general proficiency together with language specific features in an oral narration task (the use of null inflection and null arguments as well as article omission and misuse). Results indicated that the CLIL group performed significantly better than the matching NON-CLIL group but was equal to the older NON-CLIL group in general proficiency. Regarding the production of null arguments and inflection (see Martínez-Adrián and Gutiérrez-Mangado 2015a), no significant differences were found between the CLIL and the NON-CLIL counterparts, except for inflection, where the older NON-CLIL group performed significantly better. With respect to article omission and misuse (see Gutiérrez-Mangado and Martínez-Adrián 2018 with the same sample in Martínez-Adrián and Gutiérrez-Mangado 2015a), results showed that CLIL can aid in features that belong to the syntax-semantics-discourse interface (e.g. article use) rather than in features that belong to the syntax-morphology interface (e.g. article suppliance).

In other studies comparing CLIL to NON-CLIL learners in oral production, accuracy has been examined together with syntactic complexity. Martínez-Adrián and Gutiérrez-Mangado (2015b) investigated the relationship between L1 use and morphosyntactic features (word order, production of the definite and indefinite article, production of simple and complex sentences and variety of tenses used) in 3rd year CLIL and NON-CLIL learners who had received 910 and 792 h of exposure, respectively. The analysis of the oral narration task administered revealed that despite a lower use of the L1 and a higher rate of lexical diversity and complex sentences in CLIL learners, certain inaccuracies with word order and the use of tenses were attested in these learners. Complexity measures have also been explored in written production. The comparison of CLIL to NON-CLIL learners in the 3rd and 4th year of secondary education in Lahuerta (2017) revealed a better performance in terms of sentence complexity in the case of CLIL learners.

Other studies on morphosyntactic features that have adopted a more longitudinal perspective have explored production accuracy together with syntactic complexity. Lázaro Ibarrola (2012) studied the morphosyntactic development of a CLIL and a NON-CLIL group of Basque-Spanish adolescents learning L2 English over a two year period, at Time 1, when they were 13 years old; and at Time 2, when they were 15. Despite the better results obtained by the CLIL group, the improvement observed was mainly due to higher provision rates of irregular past forms, not inflectional morphemes. She also examined the rate of subordination at both

testing times, observing a significantly higher number of subordinate sentences in the CLIL group. Lázaro Ibarrola and García Mayo (2012) focused on L1 use and the morphosyntactic development of CLIL students at two testing times: first, when they were in the 2nd year of compulsory secondary education, and second, when they were in the 4th year. Regarding morphosyntactic aspects, they examined pronominal use and verbal inflection during oral production, finding a statistical improvement in pronominal use and irregular verbs between Time 1 and Time 2, but no statistical differences as regards the regular past and the third person singular *-s*. They also looked into the rate of subordination, attesting a significant improvement in the use of subordinate sentences over time.

At present, primary school learners have begun to receive increasing attention in terms of the acquisition of morphosyntactic features (Fernández-Pena and Gallardo-del-Puerto 2021; Vraciu 2020). In particular, these studies have explored production accuracy in investigations comparing CLIL to NON-CLIL learners, specifically agreement morphology and the use of explicit subjects. Fernández-Pena and Gallardo-del-Puerto (2021) analysed agreement morphology errors and subject omission in the oral production of CLIL and NON-CLIL 11- and 12-year-old learners. At the time of testing, both groups had received 617 h of EFL instruction. CLIL learners had received 488 h of additional in-class exposure. The intergroup analysis revealed the inexistence of statistically significant differences between CLIL and NON-CLIL learners in all the features examined. However, even if differences did not reach significance, the examination of the descriptive means evinced a potential impact of CLIL instruction, such as a lower rate of placeholder *is* and null subjects. Vraciu (2020) looked into the impact of CLIL instruction on the production accuracy of suppletive and affixal verb morphology in the oral narratives of 9- and 10-year-old Catalan-Spanish bilingual learners of L3 English over the course of two academic years, when learners had received 105 and 210 in-class hours of exposure, respectively. The additional exposure received through CLIL (i.e. 1 h per week) in conjunction with EFL instruction (i.e. 3 h per week) was found to be insufficient for an increase in these learners' production accuracy in L2 English verb morphology, but it did affect the range of verb inflections employed in the picture-based narratives. A qualitative inspection of the results also showed progress in terms of affixal morphology omission and target-like use of the progressive form after two years of instruction.

The review of research findings in CLIL in Spain has uncovered the acute need to focus on primary school learners in future investigations, particularly in the case of features within the morphosyntactic domain. The study of other features related to production accuracy that pose difficulties for L1 Spanish learners deserves special attention in this population. Likewise, the review of empirical findings has unveiled the need to study syntactic complexity, as to our knowledge no studies

have been conducted in this respect. Moreover, pseudo-longitudinal studies targeting different age/proficiency groups of CLIL learners, which could shed more light on the development of linguistic properties, are non-existent in young learners. Recent studies have just focused on the comparison of CLIL to NON-CLIL learners.[2] This line of research is fully justified if we want to discover potential problematic features that deserve special attention and for which the additional meaningful exposure gained through CLIL might be insufficient.

Thus, the main aim of this paper will be to investigate production accuracy and syntactic complexity across two different age/proficiency groups of CLIL learners from the 4th and 6th grade of primary education.[3] With respect to accuracy, L1 transfer effects associated with the pro-drop parameter (i.e. subject omission and subject-verb inversion) and the third person singular *-s* morpheme omission will be explored. Syntactic complexity will be measured through the production of simple and complex clauses. This study will not only contribute to the study of specific linguistic features across age and proficiency, but will also help shed light on the most problematic ones among those tested in the case of production accuracy. To our knowledge, this type of analysis is lacking in previous investigations with young CLIL learners and morphosyntactic aspects. These are the research questions that we address in this paper:

RQ 1: Are there any differences between 4th and 6th grade learners with respect to subject omission, inversion of the subject, and third person singular *-s* omission?

RQ 2: Are there any differences between 4th and 6th grade learners with respect to the production of simple and complex sentences?

3 Methodology

3.1 Participants

The participants of this study were 26 students of a state-funded primary school located in a small town from the monolingual region of Castile and Leon in Spain.

2 Note that pseudo-longitudinal investigations with young learners are becoming increasingly more common in other research areas (i.e. Martínez-Adrián 2020; Azkarai and Imaz Agirre 2017; Iglesias-Diéguez 2020).

3 Note that the present investigation also differs from recent studies on morphosyntactic features carried out with young learners (Fernández-Pena and Gallardo-del-Puerto 2021; Vraciu 2020) in that they examine data from 5th and 6th year learners, while our study explores two non-consecutive age groups (4th and 6th year).

Table 1: Participants' information.

Group	*N*	Mean age	Age of first exposure	Total hours of exposure	Years of exposure
CLIL I	14	9.07	3	1,356	6
CLIL II	12	11	3	2,455	8

Data come from two different school years (see Table 1): CLIL I ($n = 14$), including 8 females and 6 males from the 4th year of primary education (aged 9–10); and CLIL II ($n = 12$), constituting 7 females and 5 males from the 6th year (aged 11–12). At the time of testing, all of them were studying in the CLIL section, which is implemented in this school since the 1st year of primary education, when students are aged 6. Besides, English is taught since pre-primary education for 1 h a week starting at the age of 3. Later, from the 1st year of primary education onwards, all students have three 1 h sessions of EFL instruction (i.e. 114 h per course). As for content instruction, CLIL I received one 1 h session of Arts and Crafts per week and three 1 h sessions of Science, which adds up to a total of 7 h of English exposure per week (152 h per course). On the other hand, CLIL II received three 1 h sessions of Science, one 1 h session of Arts and Crafts and 2 h of Physical Education, with a total amount of 9 h per week (228 h per course). At the moment of testing, CLIL I had received around 1,045 h of English instruction, while CLIL II had received 2,109 h.

Besides, in the CLIL I group, 13 out of 14 participants received extracurricular English classes at private language schools, with an average of 2.07 h per week and a total average of 3.96 years, which adds up to 1,356 h of English exposure. In the CLIL II group, 9 out of 12 participants received private classes, with an average of 1.94 h per week and a total average of 4.7 years, which amount to 2,455 h of exposure. Since the English level of those participants who did not attend private language school was similar to the one of those who received extra classes, they were included in the study. The proficiency level test administered at the outset of the study indicated that both groups were beginner learners, although the older group had reached a slightly better command of the language. In this respect, CLIL I learners were at the A2.1 level, and CLIL II at the A2.2 level according to the Common European Framework of Reference for Languages (CEFR; Council of Europe 2001).

In this school, CLIL teachers were non-native speakers of English and had at least a B2 level (CEFR 2001). They were primary school teachers who had been pedagogically trained to teach CLIL. The materials used, such as textbooks, were specifically designed and adapted to the CLIL programme. In addition, teachers used their own materials, as well as an online platform. Likewise, there were theme-specific activities distributed throughout the year such as ones on Thanksgiving Day, Saint Patrick's Day or April Fool's Day. CLIL lessons were characterised by the lack of tasks that draw learners' attention to form or explicit corrective feedback.

The EFL classes were also taught by a non-native teacher, although a native teacher took part in one session each week to help students with their linguistic skills. The contents of the EFL lessons were tightly linked with the ones covered in CLIL, especially vocabulary. There was also decontextualised grammar instruction as suggested in the course book.

3.2 Instruments and procedure

Once parental and school permission was issued, students completed a background questionnaire so as to collect previous information about their linguistic profile. Then, participants were tested on general proficiency by administering the listening, reading and writing parts of the Cambridge English Flyers Test (UCLES, n.d.). Subsequently, participants paired up to perform an oral narration task with visual support provided by a series of wordless pictures, which belonged to the story "Room on the Broom" (Donaldson 2012). It depicts a witch's journey with her cat. While flying on her broom, suddenly, the wind blows away her hat, wand and bow. However, a dog, a frog and a bird find her objects and they all ride on her broom until it breaks in two. Then, they are threatened by a hungry dragon who wants to eat the witch but the animals come up with a plan to save their human friend: they all cover in mud and scare the dragon. Finally, they join to prepare a magic spell in the witch's cauldron, where she creates a new broom on which they can fly together.

This story was selected due to the vocabulary used, which was found to be appropriate for 4th and 6th grade students; the colourfulness of the pictures to call the participants' attention; and the number of characters that took part in the story, which would prompt the use of personal pronouns, as well as the third person singular *-s* morpheme to describe each of the character's actions. Note also that this type of tasks has been used in a wide range of studies with primary and secondary school learners in CLIL and NON-CLIL settings (see Martínez-Adrián 2020; Martínez-Adrián and Gutiérrez-Mangado 2009, 2015a, 2015b; Gutiérrez-Mangado and Martínez-Adrián 2018; Gallardo-del-Puerto and Gómez Lacabex 2013; García Mayo and Lázaro Ibarrola 2015; García Mayo and Villarreal Olaizola 2010; Lázaro Ibarrola 2012).

Data were gathered in two different sessions, except for the background questionnaire, which they were asked to complete at home with their parents' help. During the first session, the two CLIL groups took the Flyers Test individually in their respective classrooms. They were told that these results would not affect their English marks whatsoever and they were provided with precise instructions for each part. They were first asked to perform the listening part and later, they completed the reading and writing part. The scores obtained in this test were used

to match the participants in pairs for the second task. In total, there were thirteen dyads. During the second session, each proficiency-matched dyad carried out the oral narration task in a separate and quiet room: each participant was given four vignettes which they had to describe individually. Later, in pairs, the participants had to order the pictures and tell the researcher the story depicted in them. On average, the dyadic-interaction in the CLIL I group lasted for 9 min and 4 s, while in the CLIL II group lasted for 6 min and 59 s. All their production was orthographically transcribed and codified in CHILDES format (MacWhinney 2000).

3.3 Data analysis

The same procedure for data coding employed in previous studies on the acquisition of morphosyntactic features by CLIL learners was followed (Martínez-Adrián and Gutiérrez-Mangado 2009, 2015a, 2015b). The oral data were coded in terms of subject omissions, subject-verb inversions, third person singular *-s* omissions and syntactic complexity.

3.3.1 Subject omission

Obligatory contexts for overt subjects were computed in order to obtain the percentage of subject omissions per participant. An example of subject omission is given in (4). However, it should be noted that the omission of the subject in coordinated sentences was not counted as an error, as in (5) and (6):

(4) Are in a *bosque* [“forest”] (CLIL II student 05)

(5) The frog is happy with *varita* [“wand”] and jump in the water (CLIL II subject 08)

(6) It’s in the mountains and have a *lacito* [“little bow”] (CLIL I subject 05)

Neither was it taken into account after a brief hesitation to think, as in (7):

(7) The witch eh … forgot her *varit* [“wand”] (CLIL II subject 01)

3.3.2 Subject-verb inversion

For subject-verb inversion, out of the total number of subjects produced by each participant, the percentage of subject-verb inversions was obtained. Examples of inversions are shown in (8) and (9):

(8) Fells down the *varity* [“wand”] (CLIL II student 02)

(9) Then fells down the hat of the witch (CLIL II student 01)

3.3.3 Third person singular *-s* omission

As regards the third person singular *-s* morpheme, all the obligatory contexts for lexical verbs were identified. Later, the missing third person singular *-s* morphemes were quantified. Then, the percentage of omissions was obtained. An example of a third person singular *-s* omission is shown in (10):

(10) And the witch find the *varita* ["wand"] (CLIL II student 04)

Nevertheless, it should be noted that neither the use of a verb in its base form preceded by the placeholder *is* nor the plural contexts were taken into account, as in examples (11) and (12), respectively:

(11) The witch is put the … the hat (CLIL II student 08)

(12) And the witch and the hat fly (CLIL II student 07)

Besides, the incorrect use of verbs such as *be* or *have* and the use of anglicised Spanish verbs were not quantified either, as in examples (13) and (14), respectively:

(13) The witch have a hat (CLIL I student 04)

(14) The dog *apareis* ["appears"] with the cat (CLIL I student 03)

3.3.4 Syntactic complexity

Regarding the codification of syntactic complexity, first the total number of sentences produced by each participant was analysed. Later, the number of simple and complex sentences was also examined in order to obtain the percentage of simple and complex sentences, respectively. Finally, complex clauses were further subdivided into causal, relative, *that*-clause, time clause, infinitival clause and *if*-clause, and the percentage of each type of complex sentences was obtained. Examples of complex sentences are shown in (15) (*that*-clause), (16) (relative clause) and (17) (causal clause):

(15) I think that it is the first (CLIL II student 02)

(16) This is a witch that are in a *escoba* ["broom"] (CLIL II student 05)

(17) The second is this because he go to … (CLIL II student 02)

3.4 Statistical analyses

As for the statistical analyses, results were analysed in SPSS 24 (IBM Corp. 2016). Descriptive and inferential analyses were conducted. Both means and standard deviations were calculated. As for inferential analyses, the Saphiro–Wilk test was used in order to check for normality of distribution of the data. In the cases where the data was normally distributed, *T*-tests were used for intergroup analyses. When normality was not obtained, the Mann–Whitney U test was conducted. Intragroup analyses were also carried out so as to identify the most problematic features among the ones examined for production accuracy. In this respect, the Friedman test and the Wilcoxon Signed Rank test were performed. Statistical significance was indicated at <0.05 (*) and <0.01 (**) levels, and at <0.09 (#) for marginal differences (i.e. statistical tendencies).[4] Cohen's effect size values were also calculated. Following the specific benchmarks for L2 acquisition proposed by Plonsky and Oswald (2014) for intergroup comparisons, *d* values around 0.40 were considered 'small'; 'medium' if about 0.70; and 'large' if above 1.00. For intragroup contrasts, this new scale considers a *d* value of 0.60 as 'small', 1.00 as 'medium' and 1.40 as 'large'.

4 Results

In this section, we will show the results of the analyses performed to find answers to the two research questions addressed in the paper. Tables 2–4 present the intergroup analyses conducted to explore the differences between both groups in terms of production accuracy measures (subject omission, inversion of the subject and third person singular *-s* omission) (RQ1) and Tables 9 and 10 for syntactic complexity (production of simple and complex sentences) (RQ2). Intragroup analyses are also offered in Tables 5–8 so as to identify the most problematic features among the ones examined and in Figures 1 and 2 to explore the distribution of simple and complex clauses in each group.

4.1 Accuracy measures

4.1.1 Subject omission

In order to identify the existence of any differences between both groups in subject omission, omission rates were calculated as a percentage of all obligatory contexts.

4 Taking into account the scarcity of research with young learners on specific morphosyntactic features, marginal differences might reveal tendencies that could be further explored in a larger sample.

Table 2: Mean percent of subject omissions, standard deviations in parentheses and *T*-test results by learner group.

Group	Mean% subject omissions (SD)	*T*-test		Effect size
		T	*p*-value	
CLIL I	20.40 (20.16)	–1.211	0.238	$d = 0.476$
CLIL II	12.81 (8.51)			

Table 2 presents the mean percent of subject omissions, the standard deviation and the *T*-test results by learner group. As can be observed in Table 2, subject omissions were more common in CLIL I than in CLIL II. However, the intergroup comparison did not reveal statistically significant differences ($t = -1.211$, $p = 0.238$, $d = 0.476$). A descriptive inspection of the results indicated that in CLIL I, 10 out of 14 (71.42%) learners produced null subjects. The production of null subjects in this group ranged from 10.52% to 66.67%. In the CLIL II group, 10 out of 12 learners (83.33%) produced null subjects and their production of null subjects ranged from 8.33% to 30%.

4.1.2 Subject-verb inversion

In order to investigate the existence of any differences between CLIL I and CLIL II in subject-verb inversion, subject-inversion rates were calculated as a percentage of the total number of sentences with explicit subjects.

The mean percent of subject-verb inversions, standard deviations and Mann–Whitney U test results are shown in Table 3. As can be observed, subject-inversion is higher in the CLIL II than in the CLIL I group. Nevertheless, the Mann–Whitney U test did not yield significant differences between both groups ($z = -1.495$, $p = 0.135$, $d = 0.412$).

Table 3: Mean percent of subject-verb inversions, standard deviations in parentheses and Mann–Whitney U test results by learner group.

Group	Mean% subject-verb inversions (SD)	Mann–Whitney U		Effect size
		Z	*p*-value	
CLIL I	0.89 (3.34)	–1.495	0.135	$d = 0.412$
CLIL II	1.73 (2.96)			

4.1.3 Third person singular *-s* omission

So as to explore the existence of any differences in the third person singular *-s* omission between both groups, omission rates of this morpheme were calculated as a percentage of the obligatory contexts of third person singular *-s* forms.

The mean percent of subject omissions, standard deviations and Mann–Whitney U test results are shown in Table 4. The CLIL II group omitted the third person singular *-s* to a higher extent than their younger and less proficient counterparts. However, this difference did not reach significance ($z = -0.689$, $p = 0.56$, $d = 0.244$).

Table 4: Mean percent of third person singular *-s* omissions, standard deviations in parentheses and Mann–Whitney U test results by learner group.

Group	Mean% third person singular *-s* omissions (SD)	Mann–Whitney U		Effect size
		Z	*p*-value	
CLIL I	61.90 (48.66)	−0.689	0.56	$d = 0.244$
CLIL II	78.69 (29.56)			

On the whole, the intergroup analysis of the data examined could not reveal statistically significant differences for accuracy measures: subject omission, subject-verb inversion and third person singular *-s* omission.

4.1.4 Comparison of subject omission, subject-verb inversion and third person singular *-s* omission

Table 5 presents the comparison of the three features explored in the CLIL I group. As can be observed, third person singular *-s* omissions were higher than those of subject omissions and subject-verb inversion.

In order to see whether these differences were statistically significant, a Friedman test was conducted. Statistically significant differences emerged among the categories examined ($x^2 = 11.65$, $p = 0.003$). Therefore, post-hoc analyses were

Table 5: Mean percent and standard deviation of subject omission, subject-verb inversion and third person singular *-s* omission by CLIL I.

Subject omission		Subject-verb inversion		Third person singular *-s* omission	
Mean%	SD	Mean%	SD	Mean%	SD
20.40	20.16	0.89	3.34	61.90	48.66

Table 6: Wilcoxon's Signed Rank test with paired variables.

Third person singular *-s* omission versus subject omission			Third person singular *-s* omission versus subject-verb inversion			Subject omission versus subject-verb inversion		
Z	***p*-value**	**Effect size**	***Z***	***p*-value**	**Effect size**	***Z***	***p*-value**	**Effect size**
−2.344	0.019*	$d = 0.4596$	−2.807	0.005**	$d = 0.5504$	−2.670	0.008**	$d = 0.5236$

conducted. A Wilcoxon's Signed Rank test confirmed that omission of the third person singular *-s* morpheme significantly differed from subject omission ($z = -2.344$, $p = 0.019$, $d = 0.4596$) and from inversion of the subject ($z = -2.807$, $p = 0.005$, $d = 0.5504$) with small effect sizes (see Table 6).

As regards CLIL II, the same descriptive and inferential analyses were carried out (see Tables 7 and 8). Third person singular *-s* omissions were higher than those of subject omission and subject-verb inversion.

The Friedman test performed showed the existence of statistically significant differences among the categories ($x^2 = 21.83$, $p = 0.001$). Thus, post-hoc analyses were run. The Wilcoxon's Signed Rank test attested that the third person singular *-s* omission was indeed significantly higher than omission of the subject ($z = -3.059$, $p = 0.002$, $d = 0.5999$) and inversion of the subject ($z = -3.128$, $p = 0.002$, $d = 0.6134$), with small effect sizes (see Table 8).

In sum, the analysis between the three variables showed that the omission of the third person singular *-s* morpheme was the most problematic variable for both CLIL I and CLIL II groups.

Table 7: Mean percent and standard deviation of subject omission, subject-verb inversion and third person singular *-s* omission by CLIL II.

Subject omission		Subject-verb inversion		Third person singular *-s* omission	
Mean%	**SD**	**Mean%**	**SD**	**Mean%**	**SD**
12.81	8.51	1.73	2.96	78.69	29.56

Table 8: Wilcoxon's Signed Rank test with paired variables.

Third person singular *-s* omission versus subject omission			Third person singular *-s* omission versus subject-verb inversion			Subject omission versus subject-verb inversion		
Z	***p*-value**	**Effect size**	***Z***	***p*-value**	**Effect size**	***Z***	***p*-value**	**Effect size**
−3.059	0.002**	$d = 0.5999$	−3.128	0.002**	$d = 0.6134$	−2.848	0.004**	$d = 0.5585$

4.2 Complexity measures

In order to answer the second research question, which addressed the existence of differences between both age groups with respect to complexity measures, the production of simple and complex sentences was also examined. To do so, the total number of simple and complex sentences was calculated.

Table 9 illustrates the mean number of simple and complex sentences, the standard deviations and the Mann–Whitney U test results. The CLIL II group produced more instances of both simple and complex sentences than the CLIL I group. Inferential analyses yielded statistically significant differences with a large effect size when comparing both groups in terms of the production of complex sentences (z = –2.333, $p = 0.031$, $d = 0.922$), whereas regarding simple sentences, a statistical tendency was found ($z = –1.935$, $p = 0.053$, $d = 0.817$). Cohen's effect size value suggested a medium to high practical significance.

Figures 1 and 2 depict the sentence distribution of simple and complex sentences by CLIL I and CLIL II with percentages, revealing that, in both groups, the percentage of simple sentences was higher than the percentage of complex sentences.

Table 9: Mean number of simple and complex sentences, standard deviations in parentheses and Mann–Whitney U test results by learner group.

Group		Mean (SD)	Mann–Whitney U		Effect size
			Z	*p*-value	
CLIL I	Simple sentences	11.28 (7.87)	–1.935	0.053#	$d = 0.817$
CLIL II		17.25 (9.56)			
CLIL I	Complex sentences	0.5 (0.85)	–2.333	0.031*	$d = 0.922$
CLIL II		2.83 (3.51)			

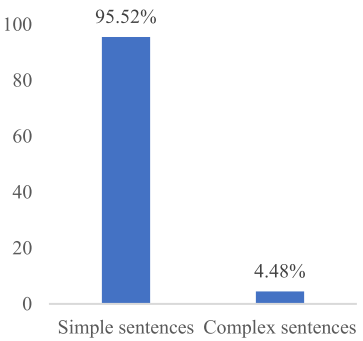

Figure 1: Sentence distribution of CLIL I.

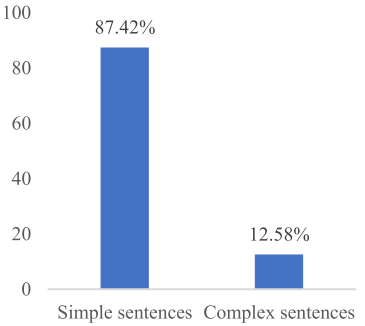

Figure 2: Sentence distribution of CLIL II.

A descriptive inspection of the data indicated that, in CLIL I, 14 out of 14 (100%) students produced simple sentences, with percentages ranging from 80% to 100%. In CLIL II, 12 out of 12 (100%) students used simple sentences, with percentages ranging from 66.67% to 100%. Regarding complex sentences, in CLIL I, only 4 out of 14 (28.57%) participants employed them, with percentages ranging from 6.06% to 20%. In the CLIL II group, 8 out of 12 (66.67%) participants resorted to complex sentences, with percentages between 8.33% and 33.33%.

Within the group of complex sentences, different types of complex sentences were identified, namely causal subordinate clauses, relative clauses, *that*-clauses, time clauses, infinitival clauses and conditional clauses.

Table 10 shows the mean of each type of complex sentences, standard deviations and Mann–Whitney U test results by learner group. The CLIL II group

Table 10: Mean of types of complex sentences, standard deviations in parentheses and Mann–Whitney U test results by learner group.

Group		Mean (SD)	Mann–Whitney U		Effect size
			Z	*p*-value	
CLIL I	Causal	0.07 (0.267)	−2.793	0.023*	*d* = 0.991
CLIL II		1.17 (1.749)			
CLIL I	Relative	1 (1.044)	−2.132	0.076#	*d* = 0.755
CLIL II		0.29 (0.726)			
CLIL I	*That*-clause	0.00 (0)	−1.080	0.742	*d* = 0.142
CLIL II		0.17 (0.577)			
CLIL I	Time	0.07 (0.267)	−1.656	0.98	*d* = 0.455
CLIL II		0.33 (0.492)			
CLIL I	Infinitival	0.07 (0.267)	−0.111	0.980	*d* = 0.02
CLIL II		0.08 (0.289)			
CLIL I	Conditional	0.00 (0)	−1.080	0.742	*d* = 142
CLIL II		0.08 (0.289)			

made use of all types of clauses and with higher means. But despite the higher production of complex sentences in the CLIL II group, no statistically significant differences were obtained when comparing both groups except for causal sentences ($z = -2.793$, $p = 0.023$, $d = 0.991$) and the statistical tendency found for relative clauses ($z = -2.132$, $p = 0.076$, $d = 0.755$). Cohen's values suggested large and medium effects, respectively.

To sum up the intergroup analysis of complexity measures, a statistical tendency was found with respect to the use of simple sentences and statistically significant differences with respect to complex sentences in favour of the older and more proficient group, particularly, in the use of causal and, arguably, relative clauses. However, despite these differences in terms of amount of sentence production, both groups preferred to use simple sentences.

5 Discussion

This section will attempt to interpret the findings for each research question. With respect to the first research question that addressed age group differences in terms of production accuracy, the results revealed no statistically significant differences between CLIL I and CLIL II regarding subject omission, subject-verb inversion and omission of the third person singular *-s*. Further, Cohen's effect size values suggested low practical significance. This finding seems to indicate that more in-class exposure to the language—obtained through the accumulated hours of EFL + CLIL instruction—does not entail that the learners perform better in these morphosyntactic features. On the whole, these findings are consistent with previous research studies examining secondary school learners that have shown that CLIL does not seem to positively affect specific areas of grammar such as overt morphology and subject production as much as general proficiency (Martínez-Adrián and Gutiérrez-Mangado 2009, 2015a, 2015b; Basterrechea and García Mayo 2013, 2014; Gutiérrez-Mangado and Martínez-Adrián 2018; García Mayo and Villarreal Olaizola 2010). Likewise, the findings of this study lend support to the ones obtained in longitudinal studies on morphosyntax with a CLIL and a NON-CLIL group, where no signs of evolution were found between both testing times in the provision of affixal morphemes (Lázaro Ibarrola 2012; Lázaro Ibarrola and García Mayo 2012). They are also in line with recent studies targeting young learners and CLIL effects (Fernández-Pena and Gallardo-del-Puerto 2021; Vraciu 2020) in that they did not find statistically significant differences between CLIL and NON-CLIL learners in terms of affixal morphology either. As these studies have shown, morphological development does not speed up until 12–13 years of age, and earlier, the additional exposure obtained through CLIL (4 h per week in CLIL I and 6 h per week in CLIL II) does not lead to faster morphosyntactic development.

The results obtained in the present study seem to align with the idea that the additional exposure obtained in CLIL is not sufficient for younger foreign language learners' implicit learning mechanisms to operate to their advantage (Housen 2012; as cited in Muñoz 2015). In this regard, a call for effective grammar pedagogy has been made in recent research (Kasprowicz and Marsden 2018). As affirmed by Roehr-Brackin (2018), in limited input environments, explicit learning is very much useful as explicit processes are fast and efficient. It is true that young learners' language analytic abilities develop at an older age (García Mayo and Villarreal Olaizola 2010), as reflected in the scant metalinguistic explanations provided by young learners when resolving Language Related Episodes during task-based interaction (see Martínez-Adrián and Arratibel-Irazusta 2020; Gallardo-del-Puerto and Martínez-Adrián In press in this respect). But it is also true that grammatical awareness can be fostered in this population, as young learners have been found not only able to draw on, but also to benefit from explicit knowledge and learning (i.e. Harley 1998; Lichtmand 2013, 2016; Milton and Alexiou 2006; as cited in Roehr-Brackin 2018). In this respect, there is a conspicuous lack of focus-on-form in CLIL classes, where corrective feedback is mainly given implicitly (Dalton Puffer 2011; Milla and García Mayo 2014) and where collaborative noticing and awareness tasks are more the exception than the norm.

It is worth noting though, that despite the inexistence of statistically significant differences between both groups in these categories, two opposing trends seem to be observed. While subject-verb omissions are more common in the CLIL I group than in the CLIL II group (see Table 2), subject-verb inversions and third person singular *-s* omissions are more productive in the CLIL II group (see Tables 3 and 4). This can be explained by the nature of their productions: the younger group produced shorter sentences, mainly used the verb *to be* and their overall productions were shorter in time. On the other hand, the CLIL II group's productions were longer and richer, so they produced more lexical verbs to describe the characters' actions, such as *fall*, *fly* or *catch*, which led them to omit more third person singular *-s* and to commit more inversion errors. The positive correlation between CLIL instruction and density of production is consistent with previous research (Gallardo-del-Puerto and Gómez Lacabex 2013; Lázaro Ibarrola 2012). The accumulated hours of EFL + CLIL instruction by age 11 could have promoted a greater production of lexical verbs in test performance (Dalton Puffer 2011) and in turn, more potential contexts for error production.

All in all, the results of the intragroup analyses also revealed that the omission of the *-s* morpheme was the most problematic variable of the three for both groups since significantly more third person singular *-s* omissions than subject omissions and subject-inversions were found. These findings are in good agreement with Slabakova's (2013) Bottleneck Hypothesis, which states that functional morphology

poses a significant problem in L2 acquisition, whereas other areas such as syntax or semantics are more easily acquired. This is also in line with previous studies on CLIL with secondary school learners that have found that, whereas syntax seems to improve as proficiency increases, morphology falls behind (Gutiérrez-Mangado and Martínez-Adrián 2018; Lázaro Ibarrola 2012). In sum, these results show that in general, learners have problems with the syntax-morphology interface, and further research should shed light on whether primary school learners also have difficulties with features pertaining to other linguistic interfaces, such as the syntax-semantics-discourse interface, as claimed by Gutiérrez-Mangado and Martínez-Adrián 2018 (2018).

As for the second research question that enquired into age group differences related to syntactic complexity, a statistical tendency was found in favour of CLIL II for simple clauses with a medium to high effect size. CLIL II learners were also found to statistically produce more complex clauses than the CLIL I group with large effect sizes. The higher number of simple sentences can be explained by the nature of the learners' productions, since the productions of the CLIL II group were longer and contained more lexical verbs than those of the less proficient group. As for complex clauses, statistically significant differences with a large effect size were found in the use of causal clauses and a statistical tendency with a medium effect size was reported in the use of relative clauses, in favour of the more proficient group (CLIL II). In fact, the descriptive inspection of the data shows learners' progression in the use of complex sentences as their proficiency increases: whereas in CLIL I, the percentages ranged from 6.66% to 20% and only 28.57% of the participants produced complex sentences; in the more proficient group, 66.67% of learners produced complex sentences, with rates ranging from 8.33% to 33.33%. This improvement in syntactic complexity was also observed in the investigations carried out with secondary school students in CLIL contexts (Martínez-Adrián and Gutiérrez-Mangado 2015b; Lahuerta 2017; Lázaro Ibarrola 2012), where the CLIL groups, who had attained a higher proficiency level, outperformed their NON-CLIL counterparts. Overall, these results suggest that syntax could truly benefit from content-based instruction and the larger exposure to the L2 entailed in these programmes (Martínez-Adrián and Gutiérrez-Mangado 2015b). Nevertheless, despite the evolution observed in CLIL II, both groups produce more simple sentences than complex sentences, which may be representative of the developmental stage they are in.

6 Conclusion

This paper has contributed to the scarcity of research on the acquisition of morphosyntactic aspects in the case of CLIL primary school learners by investigating

production accuracy and syntactic complexity across two different age/proficiency groups of CLIL learners. In particular, the comparison of both groups seems to indicate that while the accumulated exposure obtained through the combination of EFL instruction and content lessons through English by grade 6 did not so positively affect production accuracy, syntactic complexity may be enhanced by this additional exposure. The analysis of the data also showed that the omission of the third person singular *-s* morpheme was the most problematic error when compared to subject omission and inversion of the subject, confirming previous research with adolescents and adult L2 learners and lending support to Slabakova's (2013) Bottleneck Hypothesis.

In light of the results obtained, several pedagogical implications may be drawn. The implementation of focused tasks in CLIL lessons (see García Mayo 2018; Nassaji and Fotos 2011; Ranta and Lyster 2017; Shak and Gardner 2008) could boost not only the noticing of more grammar features but could also lead to more elaborated discussions about formal aspects, all of which could result in greater accuracy of not so salient linguistic features such as the third person singular *-s*. Other ventures such as stimulated recall sessions could also serve to raise learners' awareness (see the study by Bouffard and Sarkar 2008 with young learners). Likewise, the provision of more explicit feedback in CLIL lessons in the form of prompts could help in the correct identification of non-target-like forms in meaning-oriented classes (Lyster 2007; Lyster et al. 2013). The implementation of these measures together with contextualised grammar instruction in the EFL class is likely to result in more effective CLIL programmes, particularly if they are sustained from primary to secondary education. Even if the maximum benefits of CLIL might be observed in the long term as claimed in prior investigations (Nieto Moreno de Diezmas 2016; Pérez Cañado 2018; Pladevall-Ballester and Vallbona 2016), learning opportunities could be maximised through form-focused instruction already in primary education.

The results reported herein should be considered in light of certain limitations. Given the small sample size of the study, future research should aim at investigating larger groups of students. In particular, it would be convenient to do an a priori power analysis. Further studies should also address the morphosyntactic development of the same group of learners after a certain period of time in order to trace a developmental route of acquisition. The examination of a wider range of features pertaining to different linguistic interfaces would also be advisable. Likewise, further studies should also control for how grammar is specifically taught in EFL classes that CLIL students attend.

Acknowledgements: The first author would like to acknowledge the grants awarded by the Spanish Ministry of Economy and Competitiveness

(FFI2016-74950-P) (AEI/FEDER/UE), the Spanish Ministry of Science and Innovation/AEI (PID2020-113630GB-I00) and the Basque Government (IT904-16).

References

Agustín Llach, María del Pilar. 2016. Age and type of instruction (CLIL vs. traditional EFL) in lexical development. *International Journal of English Studies* 16(1). 75–96.

Azkarai, Agurtzane & Ainara Imaz Agirre. 2017. Gender and age in child interaction in an EFL CLIL context: An exploratory study. In María del Pilar García Mayo (ed.), *Learning foreign languages in primary school: Research insights*, 103–123. Clevedon: Multilingual Matters.

Basterrechea, María & María del Pilar García Mayo. 2013. Language-related episodes during collaborative tasks: A comparison of CLIL and EFL learners. In Kim McDonough & Alison Mackey (eds.), *Second language interaction in diverse educational settings*, 25–43. Amsterdam: John Benjamins.

Basterrechea, María & María del Pilar García Mayo. 2014. Dictogloss and the production of the English third person *–s* by CLIL and mainstream EFL learners: A comparative study. *International Journal of English Studies* 14(2). 77–98.

Bouffard, Laura Annie & Mela Sarkar. 2008. Training 8-year-old French immersion students in metalinguistic analysis: An innovation in form-focused pedagogy. *Language Awareness* 17(1). 3–24.

Bret Blasco, Anna. 2014. *L2 English young learners' oral production skills in CLIL and EFL settings: A longitudinal study*. Barcelona: Universitat Autònoma de Barcelona PhD Dissertation. Available at: http://hdl.handle.net/10803/285613.

Celce-Murcia, Marianne. 1991. *Teaching English as a second or foreign language*. Boston, MA: Heinle & Heinle.

Council of Europe. 2001. *Common European Framework of Reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

Dalton-Puffer, Christiane. 2011. Content and language integrated learning: From practice to principles? *Annual Review of Applied Linguistics* 31. 182–204.

Dalton-Puffer, Christiane & Ute Smit. 2013. Content and Language integrated learning: A research agenda. *Language Teaching* 46(4). 545–559.

Donaldson, Julia. 2012. *Room on the broom*. London: Penguin Books.

Fernández-Pena, Yolanda & Francisco Gallardo-del-Puerto. 2021. Agreement morphology errors and null subjects in young (non-)CLIL learners. *Vigo International Journal of Applied Linguistics* 18. 59–95.

Gallardo-del-Puerto, Francisco & Esther Gómez-Lacabex. 2013. The impact of additional CLIL exposure on oral English production. *Journal of English Studies* 11. 113–131.

Gallardo-del-Puerto, Francisco & Martínez-Adrián. In press. Task modality effects on young learners' language related episodes in collaborative dialogue. *RESLA*.

García Mayo, María del Pilar. 2003. Age, length of exposure and grammaticality judgements in the acquisition of English as a foreign language. In María del Pilar García Mayo & María Luisa García Lecumberri (eds.), *Age and the acquisition of English as a foreign language*, 94–114. Clevedon: Multilingual Matters.

García Mayo, María del Pilar. 2018. Child task-supported interaction in the Spanish EFL setting. Research and challenges. *International Journal of English Studies* 18(2). 119–143.

García Mayo, María del Pilar & Ainara Imaz Agirre. 2019. Task modality and pair formation method: Their impact on patterns of interaction and LREs among EFL primary school children. *System* 80. 165–175.

García Mayo, María del Pilar & Amparo Lázaro Ibarrola. 2015. Do children negotiate for meaning in task-based interaction? Evidence from CLIL and EFL settings. *System* 54. 40–54.

García Mayo, María del Pilar & Izaskun Villarreal Olaizola. 2010. The development of suppletive and affixal tense and agreement morphemes in the L3 English of Basque–Spanish bilinguals. *Second Language Research* 27(1). 129–149.

Gutiérrez-Mangado, M. Juncal & María Martínez-Adrián. 2018. CLIL at the linguistic interfaces. *Journal of Immersion and Content-Based Language Education* 6(1). 85–112.

IBM Corp. 2016. *IBM SPSS Statistics for Windows, Version 24.0*. Armonk, NY: IBM Corp.

Iglesias Diéguez, Kevin. 2020. *Are vocabulary tasks effective for vocabulary acquisition? Evidence from collaborative dialogue in young EFL learners*. Vitoria-Gasteiz: The University of the Basque Country Unpublished MA Dissertation.

Jiménez Catalán, Rosa María & Julieta Ojeda Alba. 2009. Girls' and boys' lexical availability in EFL. *International Journal of Applied Linguistics* 158(1). 57–76.

Jiménez Catalán, Rosa María, Yolanda Ruiz de Zarobe & Jasone Cenoz. 2006. Vocabulary profiles of English foreign language learners in English as a subject and as a vehicular language. *Vienna English Working Papers (VIEWS)* 15(3). 23–27.

Kasprowicz, Rowena E. & Emma Marsden. 2018. Towards ecological validity in research into input-based practice: Form spotting can be as beneficial as form–meaning practice. *Applied Linguistics* 39. 886–911.

Lahuerta, Ana Cristina. 2017. Syntactic complexity in secondary-level English writing: Differences among writers enrolled on bilingual and non-bilingual programmes. *Porta Linguarum* 28. 67–80.

Lasagabaster, David. 2008. Foreign language competence in content and language integrated courses. *The Open Applied Linguistics Journal* 1. 30–41.

Lázaro Ibarrola, Amparo. 2012. Faster and further morphosyntactic development of CLIL vs. EFL Basque-Spanish bilinguals learning English in high-school. *International Journal of English Studies* 12(1). 79–96.

Lázaro Ibarrola, Amparo & María del Pilar García Mayo. 2012. L1 use and morphosyntactic development in the oral production of EFL learners in a CLIL context. *International Review of Applied Linguistics* 50. 135–160.

Lyster, Roy. 2007. *Learning and teaching languages through content: A counterbalanced approach*. Amsterdam: John Benjamins.

Lyster, Roy, Kazuya Saito & Masatoshi Sato. 2013. Oral corrective feedback in second language classrooms. *Language Teaching* 46(1). 1–40.

MacWhinney, Brian. 2000. *The CHILDES project: Tools for analyzing talk*, 3rd edn. Mahwah, NJ: Lawrence Erlbaum Associates.

Marsh, David & John Masih. 1996. Teaching content through a foreign language in Finland. In Gianna Fuhauf, Do Coyle & Ingeborg Christ (eds.), *European models of bilingual/immersion education: European platform for Dutch education*, 45–67. The Netherlands: Den Haag.

Martínez-Adrián, María. 2020. ¿Los juntamos? A study of L1 use in interactional strategies in CLIL and NON-CLIL learners. *IRAL, International Review of Applied Linguistics* 58(1). 1–27.

Martínez-Adrián, María & M. Juncal Gutiérrez-Mangado. 2009. The acquisition of English syntax by CLIL learners in the Basque Country. In Yolanda Ruiz de Zarobe & Rosa María Jiménez Catalán (eds.), *Content and language integrated learning: Evidence from research in Europe*, 176–196. Bristol: Multilingual Matters.

Martínez-Adrián, María & M. Juncal Gutiérrez-Mangado. 2015a. Is CLIL instruction beneficial both in terms of general proficiency and specific areas of grammar? *Journal of Immersion and Content-Based Language Education* 3(1). 51–76.

Martínez-Adrián, María & M. Juncal Gutiérrez-Mangado. 2015b. L1 Use, lexical richness, accuracy and complexity in CLIL and non-CLIL learners. *Atlantis, Journal of the Spanish Association for Anglo-American Studies* 37(2). 175–197.

Martínez-Adrián, María, M. Juncal Gutiérrez-Mangado, Francisco Gallardo-del-Puerto & María Basterrechea. 2021. Language-related episodes by young CLIL learners. A review of task modality effects. *Language Teaching for Young Learners* 3(2). 214–246.

Martínez-Adrián, María & Izaskun Arratibel-Irazusta. 2020. The interface between task-modality and the use of previously known languages in young CLIL English learners. *Studies in Second Language Learning and Teaching* 10(3). 473–500.

Mattheoudakis, Marina, Thomaï Alexiou & Chryssa Laskaridou. 2014. To CLIL or not to CLIL? The case of the 3rd experimental primary school in Evosmos. *Major Trends in Theoretical and Applied Linguistics* 3. 215–234.

Merino, Jon Ander & David Lasagabaster. 2018. CLIL as a way to multilingualism. *International Journal of Bilingual Education and Bilingualism* 21(1). 79–92.

Milla, Ruth & María del Pilar García Mayo. 2014. Corrective feedback episodes in oral interaction: A comparison of a CLIL and an EFL classroom. *International Journal of English Studies* 14(1). 1–20.

Muñoz, Carmen. 2015. Time and timing in CLIL: A comparative approach to language gains. In María Juan-Garau & Joana Salazar-Noguera (eds.), *Content-based language learning in multilingual educational environments*, 87–104. Springer.

Nassaji, Hossein & Sandra Fotos. 2011. *Teaching grammar in second language classrooms: Integrating form-focused instruction in communicative context*. New York: Routledge.

Navés, Teresa & Mia Victori. 2010. CLIL in Catalonia: An overview of research studies. In David Lasagabaster & Yolanda Ruiz de Zarobe (eds.), *CLIL in Spain. Implementation, results and teacher training*, 30–54. Newcastle upon Tyne: Cambridge Scholars.

Nieto Moreno de Diezmas, Esther. 2016. The impact of CLIL on the acquisition of L2 competences and skills in primary education. *International Journal of English Studies* 16(2). 81–101.

Oliver, Rhonda & Agurtzane Azkarai. 2017. Review of child second language acquisition (SLA): Examining Theories and Research. *Annual Review of Applied Linguistics* 37. 62–76.

Pérez Cañado, María Luisa. 2012. CLIL research in Europe: Past, present, and future. *International Journal of Bilingual Education and Bilingualism* 15(3). 315–341.

Pérez Cañado, María Luisa. 2018. CLIL and educational level: A longitudinal study on the impact of CLIL on language outcomes. *Porta Linguarum* 29. 51–70.

Pinter, Annamaria. 2007. Some benefits of peer-peer interaction: 10-year-old children practising with a communication task. *Language Teaching Research* 11(2). 189–207.

Pinter, Annamaria. 2011. *Children learning second languages*. Basingstoke, U.K: Palgrave Macmillan.

Pladevall-Ballester, Elisabet & Anna Vallbona. 2016. CLIL in minimal input contexts: A longitudinal study of primary school learners' receptive skills. *System* 58. 37–48.

Plonsky, Luke & Frederick L. Oswald. 2014. How big is "big"? Interpreting effect sizes in L2 research. *Language Learning* 64(4). 878–912.

Ranta, Leila & Roy Lyster. 2017. Form-focused instruction. In Peter Garrett & Josep M Cots (eds.), *The Routledge handbook of language awareness*, 40–56. Abingdon: Routledge.

Roehr-Brackin, Karen. 2018. *Metalinguistic awareness and second language acquisition*. London: Routledge.
Ruiz de Zarobe, Yolanda. 2008. CLIL and foreign language learning: A longitudinal study in the Basque Country. *International CLIL Research Journal* 1(1). 6–73.
Ruiz de Zarobe, Yolanda. 2010. Written production and CLIL: An empirical study. In Christiane Dalton-Puffer, Tarja Nikula & Ute Smit (eds.), *Language use in CLIL*, 191–212. Amsterdam: John Benjamins.
Serra, C. 2007. Assessing CLIL at primary school: A longitudinal study. *International Journal of Bilingual Education and Bilingualism* 10(5). 582–602.
Shak, Juliana & Sheena Gardner. 2008. Young learner perspectives on four focus-on-form tasks. *Language Teaching Research* 12(3). 387–408.
Shintani, Natsuko. 2012. Input-based tasks and the acquisition of vocabulary and grammar: A process-product study. *Language Teaching Research* 16(2). 253–279.
Slabakova, Roumyana. 2013. What is easy and what is hard in second language acquisition?: A generative perspective. In María del Pilar García Mayo, M. Juncal Gutierrez-Mangado & María Martínez-Adrián (eds.), *Contemporary approaches to second language acquisition*, 5–28. Amsterdam: John Benjamins.
UCLES. n.d. Cambridge English: Flyers. Available at: https://www.cambridgeenglish.org/exams-and-tests/flyers/.
Vraciu, Alexandra. 2020. Production accuracy of verb morphology in early EFL: Does primary school CLIL make a difference? *Linx* 81. 1–26.