

Lexico-Grammatical Development in Secondary Education CLIL Learners

Maria Juan-Garau, José Igor Prieto-Arranz, and Joana Salazar-Noguera

1 Introduction

Research on the impact of immersion programmes on second language acquisition has generally shown benefits in comprehension and fluency, while lexico-grammatical accuracy has traditionally been seen to lag behind (e.g. Lyster 2007). This has led scholars to postulate the need for more focus on form in immersion and semi-immersion settings (e.g. Pérez-Vidal 2007).

However, the case has also been presented for content and language integrated learning (CLIL) approaches as beneficial learning contexts for the learner to acquire and eventually master lexico-grammatical competence in the target language (TL). After all, CLIL aims at fostering the learner's overall TL competence (Dalton-Puffer 2008). Thus, as opposed to traditional Focus-on-Form formal instruction (FI), CLIL has been said to present vocabulary and grammar in “authentic”, “specific” contexts through “social activities in which students interactively construct their knowledge of language use and practices” (Wilhelmer 2008: 20–21).

This has led to claims that CLIL students cognitively process their L2 at a deeper, more intense level (Aliaga 2008). And there are even those who want to see in this approach the long-sought tool with which to bridge the gap between Krashen's (1987) desirable “acquisition” and more limited “learning”. Thus, for Coyle et al., successfully implemented CLIL involves “the subtle overlap between language learning (intentional) and language acquisition (incidental)” (2010: 11), which could lead to the effective internalisation of morphosyntactic structures.

M. Juan-Garau (✉) • J.I. Prieto-Arranz • J. Salazar-Noguera
Departament de Filologia Espanyola, Moderna i Clàssica, Universitat de les
Illes Balears (UIB), Ctra. Valldemossa Km. 7.5, 07122 Palma de Mallorca, Spain
e-mail: maria.juan@uib.eu

Apart from the mostly theoretical views highlighted above, empirical studies have also been produced providing evidence of CLIL's potential language benefits. Among these, Dalton-Puffer (2008) signals vocabulary acquisition (crucially as opposed to syntax, a field in which research has not yet noted any conclusive advantage for CLIL students, as will be seen below). In her review, Dalton-Puffer argues that TL vocabulary gains are particularly significant when lexis is dealt with explicitly in the CLIL class, in fact a common occurrence (Llinares et al. 2012: 163–172; Mesquida and Juan-Garau 2013: 126). This might be seen to back up those views (see, e.g., Pérez-Vidal 2007) advising a greater presence of Focus on Form (FoF) in the CLIL classroom.

Ruiz de Zarobe's review (2015) is largely congruent with Dalton-Puffer's (2008), adding that vocabulary gains tend to be more visible in receptive, not productive, skills. Interestingly, findings are reported that CLIL learners generally outperform their non-CLIL peers in lexical richness and sophistication, "producing a higher number of lexical inventions", which Ruiz de Zarobe interprets as evidence that CLIL may foster a higher reliance on the TL rules that may help counterbalance undesired L1 transfer which non-CLIL learners are more dependent on.

Dalton-Puffer's (2008) findings are also largely congruent with those in Aguilar and Rodríguez (2012), an impressionistic interview- and questionnaire-based study enquiring into the perceptions of a group of engineering students. Their participants perceive vocabulary growth and improved listening skills after a 15-week semester in English-medium instruction at a Spanish university. Such perceptions are very much in line with those of the 670 12–14-year-old CLIL students from eleven schools across two different English-speaking countries that Coyle (2013) reports on. Combining three different data collection methods (questionnaires, respectful discussions and LOCIT) over 1 year, Coyle finds that participants generally report improved TL vocabularies (in this case, Spanish, French or German), including "the extension of content related lexis" (2013: 256).

For their part, Jiménez Catalán and Ruiz de Zarobe (2009) have researched the receptive vocabulary of both CLIL and non-CLIL primary school students, a crucial measure which has been related to both reading comprehension and incidental word learning. Their findings show significant differences to the advantage of CLIL learners. More recently, López-González (2014) provide evidence of the usefulness of CLIL, especially intensive (i.e. not extensive) bilingual programmes, in vocabulary building among Polish secondary-school learners of Spanish. Additionally, earlier research by Jiménez Catalán et al. (2006) also showed richer, more sophisticated active vocabulary among CLIL learners, although these authors carefully avoid attributing this exclusively to CLIL. Indeed, Sylvén (2004) had already found that CLIL students in Sweden were in possession of a significantly larger vocabulary than their non-CLIL counterparts. Although CLIL may have played a role in this, it was thought that additional exposure to the English language, regardless of method or learning context, also had a role to play. Thus, Sylvén (2006) specifically enquires into the reading of English texts outside the classroom context and its possible effects on the latter. When piloting her study, Sylvén found that CLIL students were substantially more exposed to English outside school, reading English books and

checking web-based materials in English twice as much as non-CLIL students. Needless to say, this has important attitudinal and motivational implications which deserve to be studied in their own right (see, e.g., Amengual and Prieto-Arranz 2015). Surprisingly, however, in her main study Sylvén finds that “[t]he [...] extra-curricular exposure to English was [...] strikingly similar [among both groups]”, with the CLIL students showing a tendency to be more exposed to Swedish than their non-CLIL counterparts (Sylvén 2006: 50–51). This finding can of course be read in different ways, although it certainly seems to point back to the amount and quality of exposure to the English medium that CLIL offers to learners as a significant variable to be taken into account.

On the other hand, and as mentioned above, morphosyntax has been noted as one of those areas that do not particularly benefit from CLIL instruction. Thus, overall results seem to indicate that fluency tends to benefit more visibly from CLIL than accuracy, although Ruiz de Zarobe (2015) also reports “greater lexical and syntactic complexity” to be found among CLIL learners. However, she reports otherwise mixed results. For example, she states that, while research has shown that CLIL may have positive effects on some morphosyntactic aspects, many others seem to remain unaffected. This is very much in line with the results provided by García Mayo and Villarreal Olaizola (2010), showing no significant differences between CLIL and non-CLIL secondary school learners of L3 English in the Basque Country as to the acquisition of suppletive and affixal tense and agreement morphemes.

At first sight, this may be found slightly surprising when evidence has also emerged pointing to CLIL having a positive impact on the learner’s competence in other highly complex language areas. By way of example, Nikula’s (2007) report provides evidence of learners demonstrating near-native pragmatic behaviour in L2 English, although note should also be taken that her study is conducted in Finland, a country in which L2 English teaching and learning conditions might not be extrapolated to other (especially southern) European countries.

In any case, there seem to be reasons to be optimistic in the light of some of the evidence produced by the latest research. In her very vast study of CLIL student perceptions, Coyle shows that her participants (12–14-year old CLIL learners of Spanish, French or German) perceive that CLIL has aided them to “‘put together’ words into longer utterances” (2013: 256). Reporting from Hungary, Vártuki (2010) claims that CLIL students at secondary school level show higher social and academic language competence in English than their non-CLIL counterparts, the gap being particularly significant in such fields as context-appropriate lexical use, mastery of morphosyntactic rules and the discursive aspects of linguistic competence, including text coherence and adaptation to sociolinguistic context. Additionally, she puts parents’ fears at rest concluding that the generally higher linguistic performance to be found among CLIL students is not at odds with their general metalinguistic, cognitive performance. This latter result, namely that CLIL does not result in defective content processing, is also obtained by Costa and Coleman (2010) who, for their part, report from Italy in a pioneering study of Italian higher education using English as the language of instruction.

Similarly, optimistic results have been shared by Lázaro Ibarrola and García Mayo (2012). In their study, the authors highlight that it is precisely in the field of morphosyntax that their CLIL participants, Spanish secondary school students, place themselves at an advantage over their non-CLIL peers. Similar participants can be found in Lázaro Ibarrola's (2012) study of morphosyntactic development in CLIL and non-CLIL secondary-school Basque-Spanish learners of L3 English. Her results place CLIL learners at a clear advantage, with higher correction rates as to the use of inflected verbs and pronouns, and significant growth as to the use of subordination.

Other studies, however, show no significant advantage for CLIL over non-CLIL learners. This is the case of Martínez Adrián and Gutiérrez Mangado (2009), who investigate whether CLIL instruction may minimise the impact of L1 transfer on English as a foreign language (EFL) learners. Their participants, again Basque-Spanish bilinguals, were lower secondary education learners of L3 English, and their morphosyntactic competence in English was measured through oral narration. Their results show that CLIL participants only significantly outperform their non-CLIL peers in one out of four different measures. This leads the authors to conclude that, although a trend has been detected pointing to CLIL somehow contributing to the minimisation of undesired L1 transfer in L3 English, results are far from definitive.

Taking into account the existing research into the possible effects of CLIL programmes on the learner's lexico-grammatical competence, Teddick and Cammarata (2012) conclude that results thus far obtained are at best mixed. This complexity is perhaps best illustrated by Aguilar and Muñoz (2013). Reporting from Spain, the authors attempt to measure the impact of one-semester CLIL programmes at postgraduate level. Among their findings, overall improvement is detected concerning the participants' grammar skills after treatment, although this does not reach statistical significance. Their results also show a clear effect of the participants' previous TL proficiency level, with the more proficient students performing more poorly after treatment whilst the least proficient participants improve significantly after a one-semester CLIL course.

Considering, therefore, that no conclusive results have so far been obtained regarding the development of lexico-grammatical competence in CLIL contexts, and that the empirical evidence available is still scanty, it is the aim of the present study to make a contribution in this direction by presenting findings on the growth of lexico-grammatical accuracy in lower secondary education CLIL learners. Thus, we intend to find out how context of learning (CLIL and non-CLIL) affects the lexico-grammatical development in lower secondary education English learners. To fulfil this objective, the following research questions were posed:

1. How does lexico-grammatical performance develop longitudinally—over 3 years—within each context of learning (CLIL and non-CLIL)?
2. How does CLIL participants' lexico-grammatical performance compare to that of their non-CLIL counterparts when hours of exposure to the target language are equated?

2 Method

2.1 Participants

Participants were two groups (CLIL vs. non-CLIL) of 13-year-old students ($N=105$)¹ enrolled in year 2 of compulsory secondary education (CSE) at the start of the study, which coincided with the onset of the CLIL programme. They were all Catalan/Spanish bilinguals from five state-run schools in the Balearic Islands, Spain. Participants in the first group were learning either science or social science through the medium of English (CLIL group: $N=70$) in addition to English as a Foreign Language (EFL), while the informants in the second group were exclusively learning EFL (non-CLIL group: $N=35$). CLIL students had a total of 6 h of class delivered through English per week (3 h of content subjects taught in English + 3 h of EFL), whereas non-CLIL students had 3 h of EFL lessons per week. There were more male (59.7 %) than female (40.3 %) participants. Data examined in this study are part of the COLE project (see Juan-Garau and Salazar-Noguera 2015).

2.2 Research Instruments

Participants' lexico-grammatical development was analysed on the basis of their performance on a cloze test and a fill-in-the-blank tense-and-aspect test over a 3-year span.

Cloze tests are fully meaningful texts in which words have been deleted at certain intervals, so that the reader has to fill in the resulting blanks in order to reconstruct the meaning of the text (Lennon 1998). Successful cloze test completion goes beyond pure focus on form (Gibbons and Lascar 1998; Storch 1998; Keshavarz and Salimi 2007) and prompts text-level processing (Yamashita 2003), thus tapping into the learners' broader lexico-grammatical continuum, asking them to resort to their organisational knowledge. The cloze instrument used in this case included 15 gaps.

The fill-in-the-blank tense-and-aspect test used in this study contained a total of twelve blanks, which had to be filled in by marking the appropriate tense and aspect of verbs included in nine short dialogues. This type of exercise, based on the use of correct verbal forms, is mainly designed to test L2 learners' grammar skills through their ability to locate the situation at some point in time as well as to detect the internal temporal constituency of the situation (Huddleston and Pullum 2002). It is a type of task that participants were used to carrying out in their EFL classes.

¹Our study being a longitudinal one, only participants that had completed all four data collection times, as described in Sect. 2.3, were eligible for analysis. Thus, participants that dropped out or failed to complete one of the two tests under study at a given data collection time were not considered.

2.3 Procedure

CLIL and non-CLIL learners' lexico-grammatical results were examined at four data collection times corresponding to three school years: T1 (at the beginning of year 2 of CSE, when the CLIL programme started), T2 (at the end of year 2 of CSE), T3 (at the end of year 3 of CSE), and T4 (at the end of year 4 of CSE).

In order to ensure reliability, tests were piloted, administered and marked consistently. On the basis of the item analysis conducted on the pilot sample using two classical measures, the facility value and the discrimination index, certain modifications were made to the initial cloze test so as to exclude those items that had proved too difficult and had very low discrimination. Correction was led by the so-called "acceptable word" method, i.e. taking as valid not necessarily the exact missing word but any word taken as correct by the authors with the help of two experienced native English teachers. No modifications were needed in the case of the fill-in-the-blank test. Two raters were involved in test scoring. Inter-rater reliability was calculated by having 10 % of the tests scored by both raters at the start of the correction process. The concordance correlation coefficient revealed a very strong agreement (0.98) between them. The few existing disagreements were discussed and settled before the remaining tests were assessed. To guarantee the requirements of validity, in the development of both tests, care was taken to include items that were deemed to measure the lexico-grammatical competence acquired by students through either context of learning (FI or CLIL).

The following statistical analyses were applied. After conducting satisfactory Kolmogorov-Smirnov and Shapiro-Wilk normality tests on our sample, the mean scores obtained at each data collection time (T1, T2, T3 and T4) for each of the measures, cloze and tense and aspect, were first compared using ANOVA tests and then by means of paired comparisons conducted with the Tukey technique. Intra-group analyses and inter-group analyses were carried out. Regarding inter-group analyses, both groups were compared by keeping the hours of exposure to the target language constant. Thus, CLIL participants at T2 (end of year 2 of CSE; age 14) were compared to their non-CLIL counterparts at T3 (end of year 3 of CSE; age 15).

3 Results

Results corresponding to participants' longitudinal lexico-grammatical development in CLIL and non-CLIL learning contexts are presented next for the cloze and tense-and-aspect tests. Additionally, comparisons between these two groups' performance in the aforementioned tests are provided.

Table 1 Cloze descriptive statistics for CLIL participants at T1, T2, T3 and T4

	<i>N</i>	Minimum	Maximum	Mean	SD
T1 cloze	70	0	10	4.4571	3.0923
T2 cloze	70	0	13	6.9286	3.4489
T3 cloze	70	0	14	7.8714	3.3120
T4 cloze	70	1	15	9.5714	3.2845

SD standard deviation

Table 2 Cloze statistics for non-CLIL participants at T1, T2, T3 and T4

	<i>N</i>	Minimum	Maximum	Mean	SD
T1 cloze	35	0	9	3.5714	2.0881
T2 cloze	35	0	11	4.3143	2.8754
T3 cloze	35	0	13	6.0571	3.6051
T4 cloze	35	1	15	7.5429	3.5930

SD standard deviation

3.1 Cloze

3.1.1 CLIL Group

Basic statistical information corresponding to cloze test scores for the CLIL group on a 15-point scale at T1, T2, T3 and T4 can be found in Table 1. The mean column indicates that CLIL participants make steady progress in the lexico-grammatical domain over the period under scrutiny.

A one-way within-subjects ANOVA with four levels (T1-T2-T3-T4) was applied to cloze measures revealing significant differences between data collection times for CLIL learners. Post-hoc paired comparisons were subsequently carried out using Tukey tests. Such paired comparisons produced significant differences between T1-T2 ($p < 0.000$), with a 2.471 increase, T3-T4 ($p = 0.014$), with a 1.700 rise, and T1-T4 ($p < 0.000$), with an overall 5.114 increment, while the growth detected between T2-T3 did not reach significance ($p = 0.334$). These results suggest that combined CLIL and EFL instruction had a positive effect on CLIL participants leading to visible overall gains, as well as gains in two of the three academic years considered.

3.1.2 Non-CLIL Group

The descriptive statistics corresponding to cloze test scores for the non-CLIL group at the different data collection times are presented in Table 2. Similarly to what has been observed in relation to the CLIL group, the mean column shows that there is a tendency towards progressive improvement for non-CLIL participants over the period studied.

As in the case of the CLIL group, data were submitted to an ANOVA analysis that revealed significant differences between times. Post-hoc Tukey comparisons between T1-T2, T2-T3 and T3-T4 produced no significant differences regarding the means obtained by non-CLIL participants ($p=0.753$, $p=0.096$, and $p=0.198$, respectively), indicating that the progress observed did not reach significance in cloze scores after any single academic year of EFL instruction. Significant differences, however, appeared after 2-year spans and overall: between T1 and T3 ($p<0.006$), with a 2.486 increase, between T2 and T4 ($p<0.000$), with a 3.229 rise, and between T1 and T4 ($p<0.000$), with a 3.971 global increase.

3.2 Tense and Aspect

3.2.1 CLIL Group

Descriptive statistics for the tense-and-aspect test, on a 12-point scale, are provided in Table 3 below. As was the case with cloze test analyses, results reveal a tendency for CLIL learners to gradually improve performance as regards the target-like use of tense and aspect forms.

CLIL participants' tense-and-aspect results were analysed through a one-way within-subjects ANOVA, with performance in the test as the dependent variable and time as the independent variable, which evinced significant differences between data collection times. More specifically, post-hoc Tukey analyses revealed all comparisons, except for the T1-T2 period, to be significant (i.e. T2-T3: $p<0.000$; T3-T4: $p=0.16$; and T1-T4: $p<0.000$; with increments of 2.100, 1.114 and 4.100, respectively). These results point to overall positive effects of combined CLIL and EFL treatment for CLIL participants resulting in a more accurate use of tense and aspect in English.

3.2.2 Non-CLIL Group

Tense-and-aspect mean scores and other statistical data corresponding to the non-CLIL group are given in Table 4. Once again, data show a clear incremental trend over time.

Table 3 Tense and aspect descriptive statistics for CLIL participants at T1, T2, T3 and T4

	<i>N</i>	Minimum	Maximum	Mean	<i>SD</i>
T1 T & A	70	0	6	1.6571	1.5108
T2 T & A	70	0	9	2.5429	2.0332
T3 T & A	70	0	9	4.6429	2.3545
T4 T & A	70	0	12	5.7571	2.6962

T & A tense and aspect, *SD* standard deviation

Table 4 Tense and aspect statistics for non-CLIL participants at T1, T2, T3 and T4

	<i>N</i>	Minimum	Maximum	Mean	SD
T1 T & A	35	0	4	1.0857	0.9061
T2 T & A	35	0	6	1.7429	1.5654
T3 T & A	35	0	9	2.8571	1.8128
T4 T & A	35	0	9	3.5714	2.3386

T & A tense and aspect, *SD* standard deviation

As in the previous sections (Sect. 3.1.1, 3.1.2, 3.2 and 3.2.1), the ANOVA analysis conducted enabled us to reject the null hypothesis. The subsequent post-hoc Tukey comparisons proved significant for 2-year periods (T1-T3: $p < 0.000$; T2-T4: $p < 0.000$) and overall (i.e. T1-T4: $p < 0.000$), but only the second academic year on its own was significant (i.e. T2-T3: $p = 0.048$) and barely so. These results suggest that EFL lessons were beneficial for the non-CLIL group in terms of increasing these learners' ability to use tense and aspect accurately in the target language. However, they needed longer than their CLIL counterparts to reap those benefits.

3.3 Comparisons Between CLIL and Non-CLIL Groups

CLIL and non-CLIL participants' results on the cloze and test-and-aspect tests were submitted to one-way between-subjects ANOVA analyses to ascertain if the performance of these two groups was significantly different at each data collection time. No significant differences were found between the two groups of learners at T1 on either test (cloze: $F = 1.424$, $df = 2, 105$, $p = 0.243$; tense and aspect: $F = 2.158$, $df = 2, 105$, $p = 0.121$), indicating that participants in the study were comparable at the start of the study in terms of their lexico-grammatical ability as shown through cloze and tense-and-aspect test completion.

By T2, differences between the groups, to the advantage of CLIL participants, were already significant in the case of the cloze test ($F = 9.067$, $df = 2, 105$, $p < 0.000$), but not yet for tense and aspect ($F = 2.116$, $df = 2, 105$, $p = 0.126$). At both T3 and T4, however, differences between CLIL and non-CLIL participants, with higher mean scores for the former, were significant for both tests (cloze T3: $F = 4.403$, $df = 2, 105$, $p = 0.015$; cloze T4: $F = 5.599$, $df = 2, 105$, $p = 0.005$; tense and aspect T3: $F = 7.678$, $df = 2, 105$, $p < 0.001$; tense and aspect T4: $F = 8.435$, $df = 2, 105$, $p < 0.000$). These results suggest that, although both groups start with comparable lexico-grammatical levels, they tend to grow apart to the advantage of the CLIL group, which seems to benefit from the CLIL programme surplus. This tendency is illustrated in Fig. 1 and particularly in Fig. 2 in relation to the cloze and tense-and-aspect tests respectively.

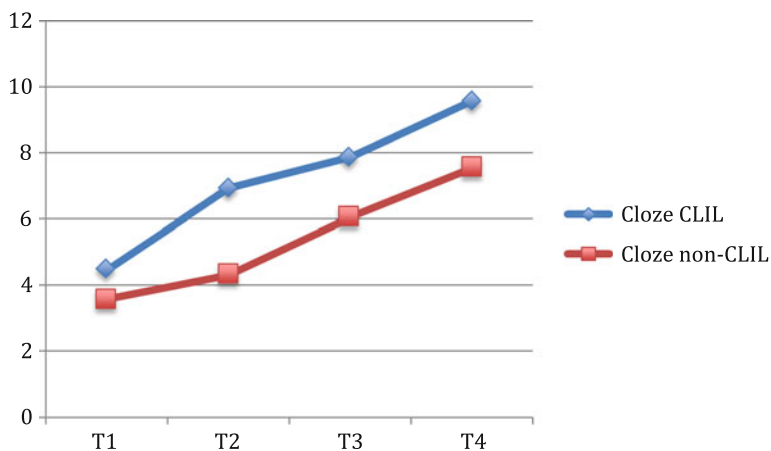


Fig. 1 Cloze: longitudinal development between T1 and T4

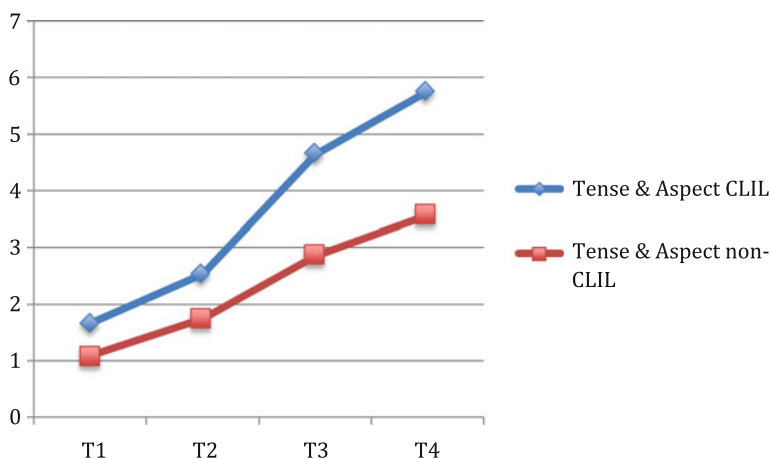


Fig. 2 Tense and aspect: longitudinal development between T1 and T4

Nonetheless, when CLIL and non-CLIL participants' performance is compared keeping hours of exposure constant (i.e. CLIL learners at T2 vs. non-CLIL learners at T3) the difference in mean scores between the two groups in both the cloze (CLIL: 6.929; non-CLIL: 6.057) and the tense-and-aspect test (CLIL: 2.543; non-CLIL 2.857) does not prove to be statistically significant. This indicates that the advantage exhibited by CLIL learners no longer holds when hours of instruction through the medium of English are the same for both groups of students.

4 Discussion

The first research question explored the extent to which lexico-grammatical performance developed over 3 years within each learning context (CLIL and non-CLIL). Results show that both CLIL and non-CLIL participants significantly improved their overall longitudinal lexical and grammatical ability. That is to say, after 3 years of instruction (T1-T4) both programmes, CLIL combined with FI and FI on its own, yielded significant differences in participants' overall achievement in the cloze and the fill-in-the-gap tense-and-aspect tests, indicating that both learning contexts appear to be beneficial for students' lexico-grammatical growth in the long term. However, our results also demonstrate that, while CLIL students significantly improved their lexico-grammatical skills each year except for cloze results between T2-T3 and the tense-and-aspect scores between T1-T2, the non-CLIL students did not significantly improve in any particular school year, apart from tense-and-aspect results between T2-T3. For the latter group, significant improvement was only found after two consecutive years of FI, T1-T3 and T2-T4, for both tests. In short, significant overall longitudinal improvement was generally seen each academic year for the CLIL group, and only after 2 years for the non-CLIL group.

These results reveal that combining CLIL with FI enables students to improve their lexico-grammatical development at a faster pace than FI on its own, thereby proving that the CLIL context makes more immediate progress possible and is more effective for short-term lexical and grammatical growth. Our findings concur with the results obtained by Lázaro Ibarrola (2012), Lázaro Ibarrola and García Mayo (2012) and Vártuki (2010), who found significantly better performance by secondary education CLIL students in mastering target language morphosyntax. Our results are also in line with Bürgi's (2007) 3-year longitudinal findings from three secondary schools in Switzerland, where CLIL learners' general English proficiency and vocabulary skills were superior to their non-CLIL classmates. Similarly, Villarreal Olaizola and García Mayo's (2007) analysis of tense and agreement inflectional morphology in oral English yielded significantly better end results from CLIL secondary students in the use of the third person singular *-s* verb marker. Hüttner and Rieder-Bünemann's (2007) results also pointed to the pre-eminence of CLIL secondary school students' skills in some micro-level features, such as consistency in the use of tenses and correct use of verbal forms.

Our results also support the findings by Dalton-Puffer (2008), Coyle (2013) and López-González (2014) on CLIL secondary education students' vocabulary growth as well as those in primary education scenarios by Jiménez Catalán et al. (2006) and Jiménez Catalán and Ruiz de Zarobe (2009), who reported greater vocabulary acquisition in CLIL students.

Nevertheless, the superior lexico-grammatical achievement by CLIL students in our study raises the question as to whether the progress achieved by students in the CLIL group in a single school year, as opposed to 2 years in FI, is due to the additional hours in a foreign language or to the introduction of a new learning context. The question as to whether the time frame—one academic year—is possibly too

short to judge the true impact of CLIL is also posed by Muñoz (2015), who enquires how long the minimum exposure time to the target language using CLIL should be before its benefits are noticeable. For her part, Sylvén's (2006) findings reveal that both the amount and the quality of exposure to English that CLIL provides prove effective when it comes to improving learners' target language vocabulary acquisition.

The results of our study also show that the only significant lexical and grammatical growth achieved in a single school year by the FI group was in one of the three time frames assessed (T2-T3), and only in the tense-and-aspect test. This finding reveals that FI students may achieve higher levels of correction in using tense and aspect nuances, possibly due to the regular practice of these grammatical areas in the FI classroom. Nevertheless, the more complex understanding of full textual meaning required to successfully fill in cloze gaps, which goes beyond the practice of discrete language items and into discursive features, was never significantly mastered in any of the periods assessed, as revealed by cloze test results. This suggests that input-rich environments, focused on meaning over form and where L2 knowledge is usually acquired indirectly (Lantolf 2011), appear to enable higher text processing levels, empowering students to put all their formal knowledge into play and thus develop their grammar, vocabulary and reading comprehension skills. Therefore, contextual communication environments, which encourage interaction and negotiation of meaning, appear to have enabled CLIL students to incidentally acquire complex lexico-grammatical abilities. This is along the lines of Aliaga (2008), who claimed that CLIL students cognitively process L2 in a more profound manner.

The considerably regular behaviour pattern of each group (CLIL and non-CLIL), achieving significant gains after one and two school years, respectively, in two different assessment tools, indicates, to a certain extent, that these tools measure the same domain of the language—i.e. their level of lexico-grammatical accuracy in the target language.

In relation to the comparison between the two groups studied, at the start of the study (T1), both CLIL and non-CLIL learners exhibit a similar onset level of lexico-grammatical competence in English, as no significant differences appear between them in the cloze and the tense-and-aspect test at that time. Hence, the two groups are comparable as far as their initial level of lexico-grammatical competence is concerned. However, as time goes by, the difference between the CLIL and non-CLIL groups becomes significant, with the former coming out top, mainly in the cloze test, at all research times (T2, T3 and T4) and also in the tense-and-aspect test for two data collection times (T3 and T4). These results indicate that a semi-immersion communicative context which activates procedural knowledge is more advantageous than the FI context in isolation in developing students' L2 grammar skills.

The reasons for lexico-grammatical growth by CLIL participants may relate to the type of test used. While fill-in-the-gap tense-and-aspect exercises are not unusual in the FI setting, cloze tests are more holistic and thus more complex, as students have to look beyond the gap's immediate context to fill in each blank with a suitable word, which involves making use of one's lexico-grammatical knowledge in a textual context.

Hence, in the case of the cloze, the CLIL setting, which is more linguistically demanding, appears to enhance the students' overall lexico-grammatical accuracy.

The second research question explored how CLIL students' lexico-grammatical performance compared to that of their non-CLIL counterparts when hours of exposure to the target language were equated. A comparison of CLIL students at T2 (end of year 2 of CSE; age 14) and their non-CLIL counterparts at T3 (end of year 3 of CSE; age 15), when hours of exposure were kept constant, found no significant differences between these two groups in either the cloze or the tense-and-aspect test. Thus, when accumulated hours of foreign language instruction are the same, the CLIL group does not obtain better results than the FI group, implying that the additional hours were beneficial to CLIL students but did not grant them a clear advantage in lexico-grammatical competence over their non-CLIL peers. The former students, who were 1 year younger and possibly had lower cognitive development but certainly more exposure, could acquire the same target language developmental level as the FI group. On the one hand, it can be interpreted that what CLIL participants learn in a formal EFL context may then be transferred to a context with added practical content (DeKeyser 2007) and, on the other hand, that students may benefit from a semi-immersion context as long as they are developmentally ready to acquire given linguistic forms (Ellis 2005).

However, our findings from the older non-CLIL students obtaining the same results as younger CLIL learners can also be interpreted in line with other scholars (e.g. Villarreal Olaizola 2011; Muñoz 2015) who claim that, with higher cognitive development but lower exposure to the target language, good results in lexico-grammatical accuracy can also be achieved through FI.

Finally, subject specialists' insufficient L2 proficiency (Nikula 2010; Hillyard 2011; Escobar Urmeneta 2013; Ruiz de Zarobe 2015) and limited abilities to teach through a foreign language (Whittaker and Llinares 2009), especially when explicit attention to learners' linguistic demands is required in CLIL settings (Swain 1990), might partly explain why CLIL students' lexical and grammatical development was not boosted to its maximum potential, and thus they did not do better than their older FI classmates.

Conclusions

The results of the present longitudinal study show that CLIL in combination with FI appears to accelerate lexico-grammatical learning, whereas FI on its own takes longer in order to exert the same positive effects. A significant contribution of this research is that over three consecutive years a considerably regular pattern has been found in both contexts, CLIL and FI, leading to enhanced lexico-grammatical abilities over one and two years respectively. Thus, greater target language exposure through CLIL appears to yield

(continued)

significant lexico-grammatical gains, although when the accumulated hours of instruction are equated, the superior performance of the CLIL group is attenuated. Several factors may have had an impact on CLIL and FI students attaining the same overall lexico-grammatical results. On the one hand, CLIL learners may have had an advantage due to the greater number of hours of exposure to the target language, through a semi-immersion TL learning environment whereby, upon learning a content subject through a foreign language, students become more used to inferring meaning from context and to transferring what has been learned in the EFL class to a more practical setting that focuses on meaning. However, this progress could be offset by the scarce response to explicit formal questions arising from semi-immersion environments, and by the lower cognitive development of the younger CLIL students. On the other hand, for non-CLIL students, greater cognitive development and enhanced practice in EFL settings of exercises focused exclusively on linguistic form may have had a positive bearing.

The question our study thereby raises is, given the significant results obtained in the development of lexico-grammatical accuracy in a single school year for CLIL plus FI, how expedient it is to wait 2 years in order to obtain the same development through FI on its own. In order to achieve more immediate effects in the lexico-grammatical domain, the results of the present study might lead to a review of secondary education curricula in Spain as regards the number of hours per year of EFL instruction, as well as of the aims set annually as far as lexico-grammatical content and competencies to be attained through EFL sessions are concerned, avoiding a repetition of similar grammatical contents over the academic years. The question should also be considered whether more communicative activities—more focused on meaning than on grammatical accuracy—ought to be introduced into EFL sessions in a regular way, as this may promote a faster development of learners' text processing skills, which does not seem to be achieved with one single type of instruction at present. Our study demonstrates that a combination of two approaches—CLIL plus FI—may be more powerful than a single approach—FI—in order to develop overall lexico-grammatical accuracy year after year. Therefore, it would appear that a renaissance of Focus on Form is called for (Lyster 2007; Pérez-Vidal 2007; Dalton-Puffer 2009), as this intentional focus may perfectly align with incidental learning (Coyle et al. 2010) in order to maximise the linguistic opportunities provided by the CLIL environment.

Future studies should carry out more observations of CLIL and FI teaching and learning processes in order to detect the specific factors that impact lexico-grammatical development, such as the degree of explicitness involved in the formal study of the language, the actual presence of communicative activities in the classroom, and the use of learners' L1. In sum, more intensive research needs to be conducted on the CLIL and FI contexts so as to further improve the quality of foreign language teaching.

Acknowledgements We gratefully acknowledge funding from the Spanish Government (HUM2004-05442-C02-01, HUM2007-66053-C02-01/02 and FFI2013-48640-C2-2-P) and the Catalan Government (SGR2005-01086/2009-140/2014-1563).

References

- Aguilar, M., & Muñoz, C. (2013). The effect of proficiency on CLIL benefits in engineering students in Spain. *IRAL International Review of Applied Linguistics*. Article first published online: 15 January 2013. doi:10.1111/ijal.12006. <http://0-onlinelibrary.wiley.com/doi/10.1111/ijal.12006/pdf>
- Aguilar, M., & Rodríguez, R. (2012). Lecturer and student perceptions on CLIL at a Spanish university. *International Journal of Bilingual Education and Bilingualism*, 15(2), 183–197.
- Aliaga, R. (2008). Aprendizaje integrado de contenidos y lengua (AICLE). IN. *Revista Electrónica d'Investigació i Innovació Educativa i Socioeducativa*, 1, 129–138. http://www.in.uib.cat/pags/volumenes/vol1_num0/aliaga/index.html
- Amengual-Pizarro, M., & Prieto-Arranz, J. I. (2015). Exploring affective factors in L3 learning: CLIL vs. non-CLIL. In M. Juan-Garau & J. Salazar-Noguera (Eds.), *Content-based language learning in multilingual educational environments* (pp. 197–220). Berlin: Springer.
- Bürgi, H. (2007). *Im Sprachbad. Besseres English durch Immersion. Eine Evaluation zweisprachiger Ausbildungsgänge an drei kantonalen Gymnasien in der Schweiz*. Bern: HEP.
- Costa, F., & Coleman, J. (2010). Integrating content and language in higher education in Italy. *International CLIL Research Journal*, 3, 19–29.
- Coyle, D. (2013). Listening to learners: An investigation into “successful learning” across CLIL contexts. *International Journal of Bilingual Education and Bilingualism*, 16(3), 244–266.
- Coyle, D., Hood, P., & Marsh, D. (2010). *CLIL. Content and language integrated learning*. Cambridge: Cambridge University Press.
- Dalton-Puffer, C. (2008). Outcomes and processes in content and language integrated learning (CLIL): Current research from Europe. In W. Delanoy & L. Wolkman (Eds.), *Future perspectives for English language teaching* (pp. 139–157). Heidelberg: Carl Winter.
- Dalton-Puffer, C. (2009). Communicative competence and the CLIL lesson. In Y. Ruiz de Zarobe & R. M. Jiménez Catalán (Eds.), *Content and language integrated learning: Evidence from research in Europe* (pp. 197–214). Bristol: Multilingual Matters.
- DeKeyser, R. (2007). Skill acquisition theory. In J. Williams & B. VanPatten (Eds.), *Theories in second language acquisition: An introduction* (pp. 97–113). Mahwah: Lawrence Erlbaum.
- Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition*, 27, 141–172.
- Escobar Urmeneta, C. (2013). Learning to become a CLIL teacher: Teaching, reflection and professional development. *International Journal of Bilingual Education and Bilingualism* (special issue), 16(3), 334–353.
- García Mayo, M. P., & Villarreal Olaizola, I. (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.
- Gibbons, J., & Lascar, E. (1998). Operationalising academic language proficiency in bilingual research. *Journal of Multilingual and Multicultural Development*, 19(1), 40–50.
- Hillyard, S. (2011). First steps in CLIL: Training the teachers. *Latin American Journal of Content & Language Integrated Learning*, 4(2), 1–12.
- Huddleston, R. D., & Pullum, G. K. (2002). *The Cambridge grammar of the English language*. Cambridge: CUP.
- Hüttner, J., & Rieder-Bünemann, A. (2007). The effect of CLIL instruction on children’s narrative competence. *Vienna English Working Papers (VIEWS)*, 16(3), 20–27.

- Jiménez Catalán, R. M., & Ruiz de Zarobe, Y. (2009). The receptive vocabulary of EFL learners in two instructional contexts: CLIL versus non-CLIL instruction. In Y. Ruiz de Zarobe & R. M. Jiménez Catalán (Eds.), *Content and language integrated learning. Evidence from research in Europe* (pp. 81–92). Bristol: Multilingual Matters.
- Jiménez Catalán, R. M., Ruiz de Zarobe, Y., & Cenoz, J. (2006). Vocabulary profiles of English foreign learners in English as a subject and as a vehicular language. *Views*, 15(3), 23–27.
- Juan-Garau, M., & Salazar-Noguera, J. (2015). Learning English and learning through English: Insights from secondary education. In M. Juan-Garau & J. Salazar-Noguera (Eds.), *Content-based language learning in multilingual educational environments* (pp. 105–121). Berlin: Springer.
- Keshavarz, M. H., & Salimi, H. (2007). Collocational competence and cloze test competence: A study of Iranian EFL learners. *International Journal of Applied Linguistics*, 17(1), 81–92.
- Krashen, S. D. (1987). *Principles and practice in second language acquisition*. London: Prentice Hall.
- Lantolf, J. P. (2011). The socio-cultural approach to second language acquisition: Sociocultural theory, second language acquisition, and artificial L2 development. In D. Atkinson (Ed.), *Alternative approaches to second language acquisition* (pp. 24–47). London/New York: Routledge.
- Lázaro Ibarrola, A. (2012). Faster and further morphosyntactic development of CLIL vs. EFL Basque-Spanish bilinguals learning English in high-school. *IJES International Journal of English Studies*, 12(1), 79–96.
- Lázaro Ibarrola, A., & García Mayo, M. P. (2012). L1 use and morphosyntactic development in the oral production of EFL learners in a CLIL context. *IRAL International Review of Applied Linguistics*, 50(2), 135–160.
- Lennon, P. (1998). Learner-centred testing: A role for cloze? *Language Learning Journal*, 17(1), 33–40.
- Llinares, A., Morton, T., & Whittaker, R. (2012). *The roles of language in CLIL*. Cambridge: Cambridge University Press.
- López-González, A. M. (2014). The effect of instruction in Polish Spanish learners' lexical availability. In R. M. Jiménez Catalán (Ed.), *Lexical availability in English and Spanish as a second language* (pp. 139–168). Berlin: Springer.
- Lyster, R. (2007). *Learning and teaching languages through content: A counterbalanced approach*. Amsterdam/Philadelphia: John Benjamins.
- Martínez Adrián, M., & Gutiérrez Mangado, M. J. (2009). The acquisition of English syntax by CLIL learners in the Basque country. In Y. Ruiz de Zarobe & R. M. Jiménez Catalán (Eds.), *Content and language integrated learning: Evidence from research in Europe* (pp. 176–196). Bristol: Multilingual Matters.
- Mesquida, F., & Juan-Garau, M. (2013). CLIL instruction and its effects on the development of negotiation strategies. *Miscelánea: A Journal of English and American Studies*, 47, 125–144.
- Muñoz, C. (2015). Time and timing in CLIL: A comparative approach to language gains. In M. Juan-Garau & J. Salazar-Noguera (Eds.), *Content-based language learning in multilingual educational environments* (pp. 87–102). Berlin: Springer.
- Nikula, T. (2007). English as an object and tool of study in classrooms: Interactional effects and pragmatic implications. In C. Dalton-Puffer & U. Smit (Eds.), *Empirical perspectives on CLIL classroom discourse* (pp. 179–204). Frankfurt: Peter Lang.
- Nikula, T. (2010). Effects of CLIL on one teacher's classroom language use. In C. Dalton-Puffer, T. Nikula, & U. Smit (Eds.), *Language use and language learning in CLIL classrooms* (pp. 105–124). Amsterdam/Philadelphia: John Benjamins.
- Pérez-Vidal, C. (2007). The need for focus on form (FoF) in content and language integrated approaches: An exploratory study. *RESLA extra 1* (monographic volume), 39–54.
- Ruiz de Zarobe, Y. (2015). The effects of implementing CLIL in education: In M. Juan-Garau, & J. Salazar-Noguera (Eds.), *Content-based language learning in multilingual educational environments* (pp. 51–68). Berlin: Springer.

- Storch, N. (1998). Comparing second language learners' attention to form across tasks. *Language Awareness*, 7(4), 176–191.
- Swain, M. (1990). Manipulating and complementing content-teaching to maximise second language learning. In R. Philipson, E. Kellerman, L. Selinker, M. Sharwood Smith, & M. Swain (Eds.), *Foreign/second language pedagogy research* (pp. 234–251). Clevedon: Multilingual Matters.
- Sylvén, L. K. (2004). *Teaching in English or English teaching? On the effects of content and language integrated learning on Swedish learners' incidental vocabulary acquisition*. Göteborg: Department of English, Göteborg University.
- Sylvén, L. K. (2006). How is extramural exposure to English among Swedish school students used in the CLIL classroom? *Views*, 15(3), 47–53.
- Teddick, D. J., & Cammarata, L. (2012). Content and language integration in K-12 contexts: Student outcomes, teacher practices, and stakeholder perspectives. *Foreign Language Annals*, 45(1), 28–53.
- Vártuki, A. (2010). Linguistic benefits of the CLIL approach: Measuring linguistic competences. *International CLIL Research Journal*, 1(3), 67–79.
- Villarreal Olaizola, I. (2011). *Tense and agreement in the non-native English of Basque-Spanish bilinguals: Content and language integrated learners vs. English as a school subject learners*. Doctoral dissertation, Universidad del País Vasco, Spain.
- Villarreal Olaizola, I., & García Mayo, M. P. (2007). *Variability in tense and agreement morphology in English as L3: CBI learners vs. ESS learners*. Paper presented at EUROSOLA 17, Newcastle-upon-Tyne, 11–14 September 2007.
- Whittaker, R., & Llinares, A. (2009). CLIL in social science classrooms: Analysis of spoken and written productions. In Y. Ruiz de Zarobe & R. M. Jiménez Catalán (Eds.), *Content and language integrated learning: Evidence from research in Europe* (pp. 215–234). Bristol: Multilingual Matters.
- Wilhelmer, N. (2008). *Content and language integrated learning (CLIL). Teaching mathematics in English*. Saarbrücken: VDM Verlag Dr. Müller.
- Yamashita, J. (2003). Processes of taking a gap-filling test: Comparison of skilled and less skilled EFL readers. *Language Testing*, 20, 267–293.