

Receptive Vocabulary of CLIL and Non-CLIL Primary and Secondary School Learners

Andrés CANGA ALONSO

Universidad de La Rioja
Departamento de Filologías Modernas
andres.canga@unirioja.es

ABSTRACT

CLIL instruction has been reported to be beneficial for foreign language vocabulary learning since CLIL students show higher vocabulary profiles than students of their same age in traditional EFL contexts. However, to our knowledge, the receptive vocabulary knowledge of CLIL and non-CLIL learners at the end of primary and secondary education has not been examined yet. Hence, this study aims at comparing the receptive vocabulary size of 79 CLIL primary learners with the receptive vocabulary knowledge of 331 non-CLIL learners at the end of primary and secondary school. Sex-based differences were also analysed. The 2k Vocabulary Levels Test (VLT) was used for the purposes of the study. Results revealed that learners' receptive vocabulary sizes lie within the most frequent 1000 words, non-CLIL secondary school students show better results than primary students but the differences between the secondary group and the CLIL group are not statistically significant. As for sex-based differences, we found no significant differences among the groups. These findings led us to believe that the CLIL approach offers a benefit for vocabulary acquisition since CLIL learners have been exposed to the foreign language for a shorter period of time and the results are quite similar to their non-CLIL secondary school partners.

Keywords: CLIL, non-CLIL, VLT, primary and secondary school, boys and girls

1. INTRODUCTION

Vocabulary is a crucial aspect in second language acquisition (SLA), overall proficiency, and general academic achievement (e.g. Daller, van Hout and Treffers-Daller 2003; Morris and Cobb 2004). In this sense, examining learners' word knowledge can provide with interesting and reliable insights into their overall language knowledge. This can be a good way of exploring the effects of CLIL instruction in the learning of English as a foreign language (EFL) by Spanish primary and secondary school students at the end of each compulsory educational level (i.e. primary and secondary school). The present paper has as its main purpose to compare the receptive vocabulary size of traditional EFL (non-CLIL) and CLIL learners at the end of primary and secondary school in Spain, and to analyse if there are statistically significant differences regarding the receptive vocabulary knowledge of male and female learners. We are interested in examining the impact of CLIL tuition on learners' receptive lexical competence versus that of general instruction in the FL.

Hence, we review the main studies dealing with receptive vocabulary size and sex-based differences. We also explore the relationship between CLIL instruction and the supposed benefits of this approach in FL vocabulary learning. Report of the study conducted with its methodology, main results found and interpretation follows. We conclude pointing out some lines for further research trying to overcome the main limitations of the present study.

2. VOCABULARY SIZE AND SEX VARIABLES IN SECOND LANGUAGE ACQUISITION

Vocabulary size is one of the key dimensions in explorations of lexical proficiency (Bulté, Housen, Pierrard and van Daele 2008). Different studies focused on finding estimations of productive and receptive vocabulary size. The tests designed for this purpose are generally built upon frequency lists (e.g. Nation 1990, 2001) on the assumption that knowledge of less frequent words implies knowledge of more frequent words (Schmitt 2000). Apart from using essays and compositions for estimations of productive vocabulary size (Laufer and Nation 1995), Laufer and Nation (1999) conceived a productive version of the Vocabulary Levels Test (VLT), which required learners to complete given sentences with the target word. These tests allow for profiling learners' vocabulary size on the basis of the frequency level to which the words used belong.

The instruments designed to test receptive vocabulary size are more numerous, e.g. Yes/No test (Meara and Buxton 1987), Eurocentres Vocabulary Size Test (Meara and Jones 1990), the European DIALANG system, P_Lex (Meara and Bell 2001), or Lex30 (Meara and Fitzpatrick 2000), just to mention some of the most frequently used ones. However, the most popular of these tests of receptive vocabulary size is the Vocabulary Levels Test, VLT for short (Nation 1990, 1993, 2001; Schmitt, Schmitt and Clapham 2001). This test is made up of further subtests each corresponding to one frequency level: first thousand most frequent words (1k), second thousand most frequent (2k), third thousand most frequent (3k), fifth thousand most frequent (5k), Academic Word List (AWL), and tenth thousand most frequent (10k).

Studies that measure vocabulary size have shown that it is incremental in nature, that is, the longer the exposure to the FL, the larger the vocabulary size (Schmitt 2000; Schmitt, Schmitt and Clapham 2001; Golberg et al. 2008; Webb 2008; Agustín Llach and Terrazas Gallego 2012; Webb and Chang 2012). Estimations of vocabulary size are also related to amount of exposure and hours of instruction in the foreign language. As examples, we found the following estimations and their related hours of exposure: After around 900 hours of instruction, learners display receptive vocabulary sizes of around 1200 words (Nurweni and Read 1999), after around 1350-1500 hours, around 2000 words (Waring 1997, Horst et al. 1998), after 3000 hours, learners show

vocabulary sizes of around 4500 words (Cobb and Horst 2004). However, to our knowledge, research studies comparing Spanish students' receptive vocabulary size at the end of primary and secondary education following CLIL and traditional EFL instruction have not been conducted yet. Specifically, this study aims at comparing the receptive vocabulary knowledge of three groups of Spanish students: 6th grade CLIL primary school students, 6th grade non-CLIL primary and non-CLIL 10th graders (last year of secondary education). Students' receptive vocabulary size and sex-based differences will be also related to their ability to understand written and spoken discourse in English (Laufer 1992, Nation 1993, Nation 2001, Adolphs and Schmitt 2004, Cobb and Horst 2004).

The role of sex-based differences has also occupied an outstanding place in current research on vocabulary acquisition in CLIL and non-CLIL types of instruction. Thus, research in non-CLIL instruction concludes that boys are superior to girls in the comprehension of heard vocabulary (Boyle 1987), in a test of academic vocabulary recognition, understanding and use (Scarcella and Zimmerman 1998) and in vocabulary knowledge in the foreign language (Edelenbos and Vinjé 2000, Lynn et al. 2005). In a recent study, Canga Alonso (2013a) finds statistically significant differences in favour of boys in a study conducted with non-CLIL 10th graders (last year secondary school students).

By contrast, women perform better than men in a memorisation test of German vocabulary (Nyikos 1990). In the same vein, Jiménez Catalán and Ojeda Alba (2009) found highly significant differences in favour of females in the mean number of words produced in a lexical availability test study in which students had to respond to 15 cues.

Nevertheless, research on vocabulary acquisition and sex-based differences also purports that male and female differences and vocabulary acquisition are not always significant. Hence, Jiménez Catalán and Terrazas Gallego (2008) discover no significant sex-based differences in performance on a receptive vocabulary test implemented with non-CLIL primary students. A set of studies compiled in Jiménez Catalán (2010) also point to mixed results on gender differences and the acquisition, development, meaning and use of vocabulary by adult, adolescents, and young learners of English and Spanish in Spain, Canada, and the USA. This book relates learner gender and meaning, vocabulary use, lexical creation, lexical production, and word association, as well as how words encode both patterns of gender representation and gender identities.

In a longitudinal study on vocabulary knowledge and gender differences with non-CLIL primary and secondary students, Agustín Llach and Terrazas Gallego (2012) obtain similar results since they found very slight differences among males and females across grades in the context of Spanish primary education concerning their receptive vocabulary knowledge in non-CLIL types of instruction.

As for sex-based differences, CLIL instruction and vocabulary acquisition, to our knowledge, only Canga Alonso (2013b) correlates these three variables in a study

conducted with last year CLIL primary school students. In the same vein as Jiménez Catalán and Terrazas Gallego (2008), Jiménez Catalán (2010), Agustín Llach and Terrazas Gallego (2012) with non-CLIL learners, Canga Alonso purports that CLIL girls are slightly better than CLIL boys at the end of primary education, but these differences are non-significant. These four latter studies used the same test as the one implemented in the present study i.e. Vocabulary Levels Test (VLT) (Schmitt, Schmitt and Clapham 2001) to measure students' receptive vocabulary knowledge and were developed in the same educational background, therefore we will refer to them in following sections.

In the light of the reviewed literature on sex-based differences, we concur with Sunderland (2010) that the relationships between vocabulary and gender are not enduring, but may be context and test type-specific, being also influenced by L1, age or L2 proficiency since the studies we have previously referred to show a variety of results depending on the context (Spain and abroad), type of instruction (CLIL/non-CLIL) and even among learners when the same test is applied to students from the same learning context.

2.1. CLIL AND VOCABULARY

The acronym CLIL (Content and Language Integrated Learning) has been used as a generic term to describe all types of approaches in which a second language is used to teach certain content subjects in the curriculum other than language lessons (Dafouz and Guerini 2009, Ruiz de Zarobe 2011). The essence of CLIL is integration with a dual focus: “language learning is included in content classes (e.g. maths, history, geography [...], etc), and content from subjects is used in language learning classes” (Mehisto, Marsh and Frigols 2008: 11). CLIL also provides real and meaningful input for the learner. This input refers to the content that the teacher is presenting as well as the language for classroom management necessary to ensure that learning takes place (Muñoz 2007).

Assuming that in CLIL settings it is necessary to progress systematically in pupils' content and language learning and use, vocabulary knowledge is of paramount importance in order to favour communication in the classroom. As a result, classroom communication – interaction between peers and teachers – is at the core. There is also growing recognition that dialogic forms of pedagogy – that is, “where learners are encouraged to articulate their learning” (Coyle, Hood and Marsh 2010: 35) – are powerful tools for securing students' engagement, and understanding. What is more, CLIL teaching should focus on quality discourse understanding between students, and between learners and teachers, so that pupils have different opportunities to discuss their own learning with other peers as it progresses. Feedback is also integrated into classroom discourse to encourage interaction among apprentices. Thus, the challenge in the CLIL setting is that trainees need to engage in dialogic

interactions by using the vehicular language. As a result of this interaction, Dalton-Puffer (2007, 2008) reports that there are some areas where clear gains are observed in CLIL classrooms such as e.g. receptive skills, vocabulary, morphology, and creativity.

These areas where achievements are observed in CLIL classrooms concord with the distinction purported by Cummins' (1979, 2008) between *basic interpersonal communicative skills* (BICS) and *cognitive academic language proficiency* (CALP) as an attempt for students to catch up with their peers in academic aspects of the school language vocabulary. Thus, BICS and CALPs should be considered in order to draw educators' attention to the timelines and challenges that second language learners encounter. BICS refers to conversational fluency in a language while CALP refers to students' ability to understand and express, in both oral and written modes, concepts and ideas that are relevant to success in school, therefore they need to know vocabulary related to the content subject in order to be able to express themselves in the vehicular language.

As will be illustrated in the procedure and data gathering section, the 2,000 frequency band of Vocabulary Levels Test (VLT) is based on the frequency lists collected by West (1953), Thorndike and Lorge (1944) list, and Kucera and Francis (1967), so it measures students' ability to recognise words from these lists which relates to Cummins' concept of CALP, since learners should be able to recognise and understand the meaning of the terms provided and relate them to their definitions. Consequently, a link between CLIL instruction and vocabulary learning can be established.

As put forward above, vocabulary size grows with proficiency level and exposure to the target language (Schmitt 1998, Fan 2000, Golberg et al. 2008). Moreover, this increase follows a systematic order related to frequency, since at the lowest levels of proficiency learners are familiar with the most frequent words, but as their experience with the foreign language increases, less frequent words are incorporated into the lexicon (Barrow et al. 1999, Vermeer 2001, Milton 2009). Hence, it seems evident that a content-based approach provides more opportunities to learn either explicitly or implicitly target vocabulary in meaningful situations (Muñoz 2007, Pérez-Vidal 2009) since learners are exposed to the target language for a longer period than students' enrolled in traditional EFL classrooms. Several studies have explored this relationship between CLIL tuition and vocabulary learning. Thus, Xanthou (2011) proves that CLIL has a positive impact in a group of primary school children in Cyprus regarding students' vocabulary tests results. Her findings demonstrate that by attaching words to their surroundings, the likelihood of comprehension and retention is increased. These gains in vocabulary size are in line with other research conducted in Spain (Jiménez Catalán, Ruiz de Zarobe and Cenoz 2006; Jiménez Catalán and Ruiz de Zarobe 2009), where significant results were obtained in favour of the CLIL group in receptive vocabulary knowledge. In a similar study, Canga Alonso (2013b) found statistically significant differences between 6th grade primary students in CLIL

contexts and those enrolled in a traditional EFL approach. However, there is a lack of studies which compare the receptive vocabulary sizes of last grade primary and secondary school students in CLIL and non-CLIL settings as this study sets out to investigate.

3. METHOD

3.1. RESEARCH QUESTIONS

As it was abovementioned, the goal of the present study is to examine students' receptive vocabulary size at the end of primary and secondary school in CLIL and non-CLIL instruction. Specifically, we seek to answer the following three research questions:

1. What is the receptive vocabulary size of the students in the sample?
2. Are there significant differences between the groups according to type of instruction (CLIL vs. Non-CLIL?).
3. Can we account for significant differences regarding sex between CLIL and non-CLIL students?

3.2. PARTICIPANTS

A total of 410 students participated in the study. 132 informants were enrolled in their last year of compulsory education (10th grade /4th ESO) in a traditional non-CLIL programme. The rest of the participants were in their last year of Primary Education (6th Grade) in two different instructional contexts (CLIL and non-CLIL). Thus, 79 belonged to the CLIL approach whereas 199 followed traditional non-CLIL instruction.

The sample was homogeneous as regards social environment and mother tongue. Our sample diverges in the type of instruction students receive, i.e. CLIL vs. non-CLIL, and consequently, in the number of hours of exposure to English FL. It also varies in students' educational level (end of primary vs. end of secondary school).

Learners in the non-CLIL group are exposed to English through the English FL school subject, exclusively. However, learners in the CLIL group receive, apart from the weekly EFL lessons, input in English in the school subject Natural Sciences, which is taught through the medium of English from the first grade of primary education.

Table 1 illustrates the approximate number of hours of exposure students have received at the time of data collection:

Grade	Age	Hours of instruction
6 th non-CLIL Primary	11-12	629
6 th CLIL Primary	11-12	944
10 th Grade/4 th ESO	15-16	1049

Table 1. Hours of exposure to English FL

Despite the fact that age and hours of instruction are not then same in our sample of informants, these three groups of learners can be compared on the basis of Spanish curricula for primary and secondary school (see BOE 01/03/2014 and BOE 21/07/2007, respectively). Both documents, following the guidelines of the *Common European Framework of Reference* (CEF) (2001), point out that the main goal of learning a language in a globalised world is to understand and produce oral and written texts, according to their level of knowledge, in the foreign language. CEF associates linguistic level to six levels of reference (A1-C2). Although Spanish curricula do not correlate educational level (primary and secondary school) with CEF levels of reference, we believe that according to the hours of exposure to the FL (see table 1) CLIL primary and non-CLIL secondary students should have acquired A1⁺-A2 level. Hence, both samples of learners can be compared. Non-CLIL primary learners (A1 level) should be included to test if, as can be expected, more hours of exposure to the FL and two different types of instruction lead to higher scores in receptive vocabulary. On the other hand, in order to be able to understand spoken and written discourse in English, students will need to have a command of the 2000 most frequent English words as soon as possible (Nation 1993, Nation and Waring 1997, Webb and Chang 2012), therefore the 2kVLT is a valid and reliable instrument to measure students' receptive vocabulary knowledge as will be shown in the following section.

3.3. DATA GATHERING INSTRUMENT

The 2,000-word frequency-band (2k) from the receptive version of the Vocabulary Levels Test (VLT) was used to measure the receptive vocabulary size of the participants in this study (Schmitt, Schmitt and Clapham 2001, version 2). This test is based on the frequency lists collected by West (1953) in the General Service List and the Thorndike and Lorge (1944) list, which were checked against the list compiled by Kucera and Francis (1967). These lists which were compiled more than a half century ago are still valid and reliable to measure students' receptive vocabulary knowledge, therefore the test is not only valid and consistent in its measurements, but also it measures what it sets out to measure (Beglar and Hunt 1999; Read 2000; Schmitt, Schmitt and Clapham 2001; Zhang and Lu 2013).

In the 2k VLT (see Appendix I), test-takers have to match a target word with the corresponding definition. A total of 60 target words are used for testing. Ten groups

of six words and three definitions make up the test. Each correct answer, i.e. matching each target word with its definition is given one point, so that the maximum score of the test is 30 points.

3.4. PROCEDURES AND DATA ANALYSIS

Data were collected in one session during class time for the traditional groups and the CLIL group. The time allotted to complete the task was 10 minutes. At the beginning of the test, clear instructions together with an example were given both orally and in written form in the students' mother tongue (Spanish) to clarify what they were being asked to do.

Tests were corrected and total scores obtained. 0 was the minimum score and 30 was the maximum. Estimations in words were also obtained. In order to calculate students' word estimates, Nation's formula "Vocabulary size = N correct answers multiplied by total N words in dictionary (the relevant word list) divided by N items in test" (Nation 1990: 78) was applied.

The sample was also analysed to check whether there were statistically significant differences according to type of instruction. Thus, Kolmogorov-Smirnov statistical test was implemented to test if the sample met the normality assumption. As will be shown in the results section, our sample did not always meet this assumption, so non-parametric tests of means comparison had to be implemented. The Kruskal-Wallis rank sum was, then, chosen to compare differences between three independent groups. In the present study, the Wilcoxon test was used to measure whether students' VLT scores differed based on type of instruction and sex. The groups were compared as a whole and in pairs: 6th CLIL and non-CLIL Primary, 6th CLIL Primary and non-CLIL 10th grade, and 6th non-CLIL Primary and non-CLIL 10th grade.

4. RESULTS

Regarding our first research question (What is the receptive vocabulary size of the students in the sample?), students' scores were translated into a number of known words for each frequency level applying Nation's formula (1990: 78) abovementioned. Table 2 shows the word estimates obtained by the students involved in the study. As could be expected, 10th graders obtained the highest scores (936 words) closely followed by 6th grade CLIL students (903 words). 6th Primary non-CLIL students were far behind the other two groups analysed in the present study.

Educational Level	Word estimates
10 th Grade/4 th ESO	936
6 th CLIL Primary	903
6 th non-CLIL Primary	601

Table 2. Word estimates for CLIL and non-CLIL learners

These figures clearly indicate that the overall receptive vocabulary size of this sample of Spanish EFL learners is considerably lower than 2,000 words, regardless of their educational level.

In order to gain statistical value of the nature of the differences between the three groups (research question 2) we applied Kolmogorov-Smirnov test, and as illustrated in table 3 the p-values obtained were very low ($p < 0.05$), so non-parametric Kruskal-Wallis rank sum test was implemented.

Educational Level	Kolmogorov-Smirnov
10 th Grade/4 th ESO	$p=0.007$
6 th CLIL Primary	$p=0.04$
6 th non-CLIL Primary	$p=0.02$

Table 3. Inferential statistics

The p-value ($p < 2.2e-16$) indicates that there are statistically significant differences in the number of words these groups of learners recognised in the VLT, therefore we analysed our data as follows: 10th graders and 6th CLIL Primary, 6th CLIL Primary and 6th non-CLIL Primary and 10th graders versus 6th non-CLIL Primary.

In order to test if the differences in word estimates were significant, the Wilcoxon test was implemented. As for 10th graders and 6th Primary CLIL students the p-value ($p=0.77$) is higher than 5%, it can be concluded that the differences are not significant between last grade CLIL primary students and last grade non-CLIL secondary learners.

With regard to the differences between last grade CLIL and non-CLIL primary students, the lower p-value ($p=4.14e-12$) indicates that there are statistically significant differences between both groups of learners. As expected, our findings reveal that 10th graders are significantly better than 6th Primary non-CLIL students ($p=8.11e-15$).

As far as sex-based differences in vocabulary size are concerned (see table 4), descriptive results reveal that the highest mean was attained by boys in the 10th grade (1001 words). CLIL girls attained higher scores in the VLT than non-CLIL girls and CLIL boys and non-CLIL students at primary level. It is also outstanding that CLIL boys obtained better results than 10th grade non-CLIL girls.

Educational Level	Sex	Word estimates
10 th Grade/4 th ESO	Boys	1001
	Girls	879
6 th CLIL Primary	Boys	887
	Girls	925
6 th non-CLIL Primary	Boys	596
	Girls	606

Table 4. Word estimates for males and females

The Wilcoxon test was conducted to test inferential statistical differences among the groups. Table 5 illustrates that the differences according to sex among the groups are non-significant at a significance level of 5% in vocabulary size estimations:

	Wilcoxon Test
10 th Grade	p=0.0602
6 th CLIL Primary	p=0.9047
6 th non-CLIL Primary	p=0.9951
10 th Grade, 6 th CLIL Primary, 6 th non-CLIL Primary	p=0.62

Table 5. Results of inferential statistics for gender-based differences

5. DISCUSSION

The results for the first research question revealed that learners’ receptive vocabulary size is within the limits of the first 1,000 most frequent words. This finding is not surprising, since it is in line with previous results with primary and secondary school students in similar and diverse educational contexts (Nurweni and Read 1999, López Mezquita 2005, Agustin Llach and Terrazas Gallego 2012, Webb and Chang 2012, Canga Alonso 2013a, Canga Alonso 2013b). According to these results, it can be argued that the 1k VLT should have been implemented since the mean scores obtained by our informants in the 2k VLT are slightly below 1,000 most frequent words. However, if we had administered the 1k version of the VLT, some of the students would not have had the opportunity to show all their receptive vocabulary knowledge since 26.58% of the students in the CLIL group recognized more than 1,000 words in the test and 18.65% in the non-CLIL secondary sample. Results also throw another finding: CLIL students recognise more words in the 1100-1500 word band than our non-CLIL secondary school sample which seems to concord with Dalton-Puffer’s (2006, 2007) assertion that receptive skills (e.g. vocabulary) are favoured by CLIL instruction.

Our second main aim was to compare the receptive vocabulary sizes of learners who attend different instructional programmes at different educational levels. As expected, results point to strongly significant differences between non-CLIL students at the end of primary and secondary education in favour of the secondary group. Thus, longer FL exposure fosters vocabulary acquisition, and the benefits start

cropping up after some time (Celaya and Ruiz de Zarobe 2010). The same findings account for the CLIL and non-CLIL primary samples, which show that the longer exposure to English input CLIL learners have received can help explain this advantage in general receptive vocabulary size. Furthermore, the different nature of their exposure, in which traditional EFL instruction is combined with a more meaningful and contextualized content instruction through the L2 might also account for this difference (Xanthou 2011).

As for students' receptive vocabulary knowledge at the end of primary and secondary education after a similar amount of hours of instruction in two different types of instruction (CLIL vs. non-CLIL), our data reveal a slight non-significant difference between both groups of students. This small variance between both groups seems to show that CLIL instruction favours vocabulary learning (Dalton-Puffer 2006, 2007) since 10th graders have been exposed to the target language for one hundred hours more than the CLIL group and the results are quite alike. These data also indicate that Cummins' cognitive academic language proficiency (CALP) has been developed in the CLIL classroom since students were able to recognise and understand the meaning of the terms provided in the VLT and relate them to their definition in a similar way their non-CLIL partners did. These results led us to think that had we tested learners knowledge of academic vocabulary or specific vocabulary related to the field of science (the CLIL subject), results might have been even higher in favour of the CLIL group, whose vocabulary size might be bigger in this semantic field or frequency band. However, this is just mere speculation, since we have not conducted such test and further research which tests the words students may have acquired in the content classes is needed in order to compare their results to non-CLIL learners.

It is also outstanding that our sample of CLIL students attained higher scores than learners of a similar age who have received a similar amount of hours of CLIL instruction. Thus, after 944 hours, participants in the present study obtained 903 as their mean score of word estimates, whereas the word estimates for Jiménez Catalán and Ruiz de Zarobe's (2009) informants was 800 words after being exposed to English for a total amount of 960 hours. A more developed CLIL programme may account for the higher results attained by our informants since Jiménez Catalán and Ruiz de Zarobe's study was developed when CLIL programmes were still at the beginning of their development and implementation. However, this is just mere speculation seeing that we do not know for sure how both content programmes were introduced and implemented as we did not have the opportunity to observe how teachers perform in their CLIL classrooms. Thus, further research is called for in order to analyse how teachers interact in their CLIL classrooms to be able to assess if their performance in the classroom may or may not favour receptive vocabulary learning.

According to previous studies, these findings also show that our learners would have problems in performing certain linguistic tasks such as understanding informal

spoken discourse for which Adolphs and Schmitt (2004) estimate that, at least, 2,000 words are needed. Nation (2006) talks of 6,000 to 7,000 word families for comprehension of spoken text, if 98% coverage of a text is desired for understanding written text, and he contends that 8,000 to 9,000 word families are needed for text understanding, or for reading for pleasure. Hirsh and Nation (1992) point out that the knowledge of 5,000 word families is required to enjoy reading. Nevertheless, our sample of CLIL primary and non-CLIL secondary informants would have an easier time watching TV programmes (Webb and Rogers 2009a), or movies (Webb and Rogers 2009b), since the first 1000 most frequent words make up for around 85% of the total word coverage. Accordingly, researchers call for the command of 2,000-3,000 most frequent words as soon as possible (Nation 1993, Nation and Waring 1997, Webb and Chang 2012). Furthermore, we agree with Schmitt (2000) and Webb and Chang (2012) that such a paramount learning task cannot be left to chance and that the most frequent words should be taught explicitly in the EFL classroom.

Sex-based differences are non-significant in the present study. This finding is in line with previous research conducted in the same educational background with CLIL and non-CLIL learners (Agustín Llach and Terrazas Gallego 2012, Canga Alonso 2013b). However, it is outstanding that CLIL girls closely followed non-CLIL secondary school boys, and both CLIL boys and girls obtained higher scores in the VLT than non-CLIL secondary school girls. As abovementioned, these data seem to show that CLIL instruction has been beneficial for the CLIL primary group. However, we should take this assumption with care since the sample of CLIL learners is not very big and further research is needed in order to test if a bigger sample of CLIL students will obtain similar results to the ones shown in this paper. Finally, we agree with Sunderland's (2010) assertion that the relationships between vocabulary and sex are not enduring, but may be also influenced by age or L2 proficiency, since by implementing the same test (i.e. 2k VLT) older male non-CLIL learners obtained the highest scores whereas younger CLIL girls outperformed their male CLIL partners (i.e. same context, type of instruction and age) and our female non-CLIL secondary sample (i.e. same context, different age and type of instruction).

6. CONCLUSION

Three main findings stem from the present research study. First, the receptive vocabulary size of our Spanish primary and secondary school EFL learners lies within the range of first 1,000 most frequent words in English, according to the results of the 2k VLT. However, non-CLIL secondary learners show non-significant higher overall receptive vocabulary sizes when compared to CLIL primary students, despite having being exposed to the FL for a longer period of time. This constitutes our second main finding. Finally, no statistically significant sex-based differences have been found in the present study.

These results lead us to think that CLIL tuition with its contextualized and meaningful FL teaching is a beneficial approach for vocabulary acquisition. In this sense, further studies which explore CLIL benefits for vocabulary acquisition for longer periods of time going into the compulsory secondary education stage are needed. Additionally, this study has only been concerned with receptive vocabulary of the most frequent words; we are inclined to believe though, that productive vocabulary and lexical knowledge of less frequent words, especially from the Academic Word List can also derive great benefits from an educational approach of the content-based type. One further limitation of the present study is the use of a single, and somewhat limited, instrument to measure vocabulary size. Using other tests for vocabulary knowledge, such as lexical availability tests might throw even more insightful results and reveal more qualitative data concerning learners' vocabulary knowledge and lexical development, as well as finer differences between traditional and content-based approaches. Finally, further research can be conducted to attach receptive vocabulary knowledge to the six levels (A1-C2) of the *Common European Framework of Reference* (2001), therefore X_Lex (Meara and Milton 2003, Milton 2010) could be implemented to test CLIL and non-CLIL students' receptive vocabulary level at the end of primary and secondary school.

ACKNOWLEDGEMENTS

We would like to acknowledge the financial support of the Spanish Ministry of Science and Innovation through grant no. FFI 2010-19334. We are also very thankful to our statistician Montserrat San Martin for providing us with very valuable statistical assistance and helpful insights into data interpretation. Any remaining errors are our own.

REFERENCES

- Adolphs, Svenja and Norbert Schmitt (2004). Vocabulary coverage according to spoken discourse context. In Paul Bogaards and Batia Laufer, eds., 39-52.
- Agustín Llach, M.^a Pilar and Melania Terrazas Gallego (2012). Vocabulary knowledge development and gender differences in a second language. *ELIA* 12: 45-75. <http://institucional.us.es/revistas/elia/12/art_3.pdf> (Accessed November 6 2015).
- Arnaud, Pierre J. and Henri Béjoint, eds. (1999). *Vocabulary and Applied Linguistics*. London: Macmillan.
- Barrow, Jack; Y. Nakanishi and H. Ishino (1999). Assessing Japanese college students' vocabulary knowledge with a self-checking familiarity survey. *System* 27.2: 223-247.

- Bartning, Inge; Maisa Martin and Ineke Vedder, eds. (2010). *Communicative Proficiency and Linguistic Development: Intersections between SLA and Language Testing Research. Eurosla Monograph 1*.
- Beglar, David and Alan Hunt (1999). Revising and validating the 2000 word level and University word level vocabulary tests. *Language Testing* 16: 131-162.
- Bogaards, Paul and Batia Laufer, eds. (2004). *Vocabulary in a Second Language*. Amsterdam and Philadelphia: John Benjamins.
- Boletín Oficial del Estado (BOE 27/07/2007). Orden ECI/2220/2007 por la que se establece el currículo y se regula la ordenación de la Educación Secundaria Obligatoria. <<http://www.boe.es/boe/dias/2007/07/21/pdfs/A31680-31828.pdf>> (Accessed November 6 2015).
- Boletín Oficial del Estado (BOE 01/03/2014). Real Decreto 126/2014, de 28 de febrero, por el que se establece el currículo básico de la Educación Primaria. <<http://www.boe.es/boe/dias/2014/03/01/pdfs/BOE-A-2014-2222.pdf>> (Accessed November 6 2015).
- Boyle, Joseph P. (1987). Sex differences in listening vocabulary. *Language Learning* 37.2: 273-284.
- Bulté, Bram; Alex Housen; Michel Pierrard and Siske Van Daele (2008). Investigating lexical proficiency development over time – the case of Dutch-speaking learners of French in Brussels. *French Language Studies* 18.3: 277-298.
- Canga Alonso, Andrés (2013a). Receptive vocabulary size of secondary Spanish EFL learners. *Revista de Lingüística y Lenguas Aplicadas* 8: 67-75.
- Canga Alonso, Andrés (2013b). The Receptive Vocabulary of Spanish 6th Grade Primary School Students in CLIL instruction: A Preliminary study. *Latin American Journal of Content and Language Integrated Learning (LACLIL)* 6.2: 22-41. <doi:10.5294/laclil.2013.6.2.2> (Accessed November 6 2015).
- Celaya, M.^a Luz and Yolanda Ruiz de Zarobe (2010). First languages and age in CLIL and non-CLIL contexts. *International CLIL Research Journal* 1.3. <<http://www.icrj.eu/13/article6.html>> (Accessed November 6 2015).
- Cobb, Tom and Marlise Horst (2004). Vocabulary sizes of some City University students. *City University (HK) Journal of Language Studies* 1: 59-68.
- Council of Europe (2001). *Common European Framework of Reference for Languages*. Cambridge: Cambridge University Press.
- Coyle, Do; Philip Hood and David Marsh (2010). *CLIL: Content and Language Integrated Learning*. Cambridge: Cambridge University Press.
- Cummins, James (1979). Cognitive/academic Language Proficiency, linguistic interdependence, the optimum age question and some other matters. *Working Papers on Bilingualism* 19: 121-129.
- Cummins, James (2008). BICS and CALP: Empirical and theoretical status of the distinction. In Brian V. Street and Nancy H. Hornberger, eds., 71-83.
- Dafouz, Emma and Michael C. Guerini, eds. (2009). *CLIL Across Educational Levels*. Madrid: Richmond.

- Daller, Helmut; Roeland Van Hout and Jeanine Treffers-Daller (2003). Lexical richness in the spontaneous speech of bilinguals. *Applied Linguistics* 24.2: 197-222.
- Dalton-Puffer, Christiane (2007). *Discourse in Content and Language Integrated Learning (CLIL)*. Amsterdam and Philadelphia: John Benjamins.
- Dalton-Puffer, Christiane (2008). Outcomes and processes in Content and Language Integrated Learning (CLIL): Current research from Europe. In Werner Delanoy and Laurenz Volkmann, eds., 139-157.
- Delanoy, Werner and Laurenz Volkmann, eds. (2008). *Future Perspectives for English Language Teaching*. Heidelberg: Carl Winter.
- Edelenbos, Peter and Maria P. Vinjé (2000). The assessment of a foreign language at the end of primary (elementary) education. *Language Testing* 17.2: 144-162.
- Fan, May (2000). How big is the gap and how to narrow it? An investigation into the active and passive vocabulary knowledge of L2 learners. *RELC Journal* 31: 105-119.
- Golberg, Heather; Johanne Paradis and Martha Crago (2008). Lexical acquisition over time in minority first language children learning English as a second language. *Applied Psycholinguistics* 29: 41-65.
- Hirsh, David and Paul Nation (1992). What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language* 8: 689-696.
- Horst, Marlise; Tom Cobb and Paul Meara (1998). Beyond a clockwork orange: acquiring second language vocabulary through reading. *Reading in a Foreign Language* 11.2: 207-223.
- Jiménez Catalán, Rosa M.^a, ed. (2010). *Gender Perspectives on Vocabulary in Foreign and Second Language*. Basingstoke: Palgrave MacMillan.
- Jiménez Catalán, Rosa M.^a and Julieta Ojeda Alba (2009). Girls' and boys' lexical availability in EFL. *International Journal of Applied Linguistics* 158: 57-76.
- Jiménez Catalán, Rosa M.^a and Melania Terrazas Gallego (2008). The receptive vocabulary of English foreign language young learners. *Journal of English Studies* 5: 173-191.
- Jiménez Catalán, Rosa M.^a and Yolanda Ruiz de Zarobe, eds. (2009). *Content and Language Integrated Learning. Evidence from Research in Europe*. Bristol: Multilingual Matters.
- Jiménez Catalán, Rosa M.^a; Yolanda Ruiz de Zarobe and Jasone Cenoz (2006). Vocabulary profiles in English as a subject and as a vehicular language. *VIEWZ* 15.3: 23-27.
<http://anglistik.univie.ac.at/fileadmin/user_upload/dep_anglist/weitere_Uploads/Views/views15_3_clil_special.pdf> (Accessed November 6 2015).
- Jiménez Catalán, Rosa M.^a and Yolanda Ruiz de Zarobe (2009). The receptive vocabulary of EFL learners in two instructional contexts: CLIL versus non-CLIL. In Rosa M.^a Jiménez Catalán and Yolanda Ruiz de Zarobe, eds., 81-92.

- Kucera, Henry and Winthrop Nelson Francis (1967). *A Computational Analysis of Present Day American English*. Providence: Brown University Press.
- Laufer, Batia (1992). How much lexis is necessary for reading comprehension. In Pierre J. Arnaud and Henri Béjoint, eds., 126-132.
- Laufer, Batia and Paul Nation (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics* 16.3: 307-322.
- Laufer, Batia and Paul Nation (1999). A vocabulary-size test of controlled productive ability. *Language Testing* 16.1: 33-51.
- López-Mezquita Molina, M.^a Teresa (2005). La Evaluación de la Competencia Léxica: Tests de Vocabulario. Su Fiabilidad y Validez. Ph.D. diss., Universidad de Granada.
- Lynn, Richard; David M. Fergusson and L. John Horwood (2005). Sex differences on the WISC-R in New Zealand. *Personality and Individual Differences* 39: 103-114.
- Meara, Paul and Barbara Buxton (1987). An alternative to multiple choice vocabulary testing. *Language Testing* 4.2: 142-154.
- Meara, Paul and G. Jones (1990). *Eurocentres Vocabulary Size Tests 10KA*. Zurich: Eurocentres Learning Service.
- Meara, Paul and James Milton (2003). *X_Lex, the Swansea Levels Test*. Newbury: Express.
- Meara, Paul and Huw Bell (2001). P_Lex: A simple and effective way of describing the lexical characteristics of short L2 texts. *Prospect* 16.3: 323-337.
- Meara, Paul and Tess Fitzpatrick (2000). Lex 30: An improved method of assessing productive vocabulary in an L2. *System* 28: 19-30.
- Mehisto, Peter; David Marsh and M.^a Jesús Frigols (2008). *Uncovering CLIL: Content and Language Integrated Learning in bilingual and multilingual contexts*. Oxford: Macmillan.
- Milton, James (2009). *Measuring Second Language Vocabulary Acquisition*. Bristol: Multilingual Matters.
- Milton, James (2010): The development of vocabulary breadth across the CEFR levels. A common basis for the elaboration of language syllabuses, curriculum guidelines, examinations, and textbooks across Europe. In Inge Bartning, Maisa Martin and Ineke Vedder, eds., 211-232. <<http://eurosla.org/monographs/EM01/211-232Milton.pdf>> (Accessed November 6 2015).
- Morris, Lori and Tom Cobb (2004). Vocabulary profiles as predictors of the academic performance of teaching English as a second language trainees. *System* 32.1: 75-87.
- Muñoz, Carmen (2007). CLIL. Some thoughts on its psychological principles. *RESLA* volumen monográfico: 11-17.
- Nation, Paul (1990). *Teaching and Learning Vocabulary*. New York: Newbury.
- Nation, Paul (1993). Using dictionaries to estimate vocabulary size: Essential, but rarely followed, procedures. *Language Testing* 10: 27-40.

- Nation, Paul (2001). *Learning Vocabulary in Another Language*. New York: Cambridge University Press.
- Nation, Paul (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review / La revue canadienne des langues vivantes* 63: 59-81.
- Nation, Paul and Robert Waring (1997). Vocabulary size, text coverage and word lists. In Norbert Schmitt and Michael McCarthy, eds., 6-19.
- Nurweni, Ari and John Read (1999). The English vocabulary knowledge of Indonesian university students. *English for Specific Purposes* 18.2: 161-175.
- Nyikos, Martha (1990). Sex-related differences in adult language learning: Socialization and memory factors. *Modern Language Journal* 74: 273-287.
- Pérez-Vidal, Carmen (2009). The integration of content and language in the classroom: A European approach to education (the second time around). In Emma Dafouz and Michael C. Guerini, eds., 3-16.
- Read, John (2000). *Assessing Vocabulary*. Cambridge: Cambridge University Press.
- Ruiz de Zarobe, Yolanda (2011). Which language competencies benefit from CLIL? An insight into applied linguistics research. In Yolanda Ruiz de Zarobe, José Manuel Sierra and Francisco Gallardo del Puerto, eds., 123-153.
- Ruiz de Zarobe, Yolanda; José Manuel Sierra and Francisco Gallardo del Puerto, eds. (2011). *Content and Foreign Language Integrated Learning*. Bern: Peter Lang.
- Scarcella, Robin and Cheryl Zimmerman (1998). Academic Words and Gender. ESL Student Performance on a Test of Academic Lexicon. *Studies in Second Language Acquisition* 20: 27-49.
- Schmitt, Norbert (1998). Tracking the incremental acquisition of second language vocabulary: A longitudinal study. *Language Learning* 48: 281-317.
- Schmitt, Norbert (2000). *Vocabulary in Language Teaching*. Cambridge: Cambridge University Press.
- Schmitt, Norbert and Michael McCarthy, eds. (1997). *Vocabulary: Description, Acquisition and Pedagogy*. New York: Cambridge University Press.
- Schmitt, Norbert, Diane Schmitt and Caroline Clapham (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Level Test. *Language Testing* 18: 55-88.
- Street, Brian V. and Nancy H. Hornberger, eds. (2008). *Encyclopaedia of Language and Education, 2nd Edition, Volume 2*. New York: Springer Science + Business Media LLC.
- Sunderland, Jane (2010). Theorizing gender perspectives in foreign and second language learning. In R. M. Jiménez Catalán, ed., 1-22.
- Thorndike, Edward and Irvine Lorge (1944). *The Teacher's Word Book of 30,000 Words*. New York Teachers College: Columbia University.
- Vermeer, Anne (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics* 22: 217-34.

- Waring, Robert (1997). A comparison of the receptive and productive vocabulary sizes of some second language learners. *Immaculata* (Notre Dame Seishin University, Okayama) 1: 53-68.
- Webb, Stuart (2008). Receptive and productive vocabulary sizes of L2 learners. *Studies in Second Language Acquisition* 30.1: 79-95.
- Webb, Stuart and Michael P. H. Rodgers (2009a). The vocabulary demands of television programmes. *Language Learning* 59.2: 335-366.
- Webb, Stuart and Anna C-S Chang (2012). Second language vocabulary growth. *RELC Journal* 43.1: 113-126.
- Webb, Stuart and Michael P. H. Rodgers (2009b). The lexical coverage of movies. *Applied Linguistics* 30.3: 407-427.
- West, Michael (1953). *A General Service List of English Words*. London: Longman.
- Xanthou, Maria (2011). Current trends in L2 vocabulary learning and instruction. Is CLIL the right approach? *Advances in research on language acquisition and teaching: Selected Papers:* 459-471. <<http://www.enl.auth.gr/gala/14th/Papers/English%20papers/Xanthou.pdf>> (Accessed November 6 2015).
- Zhang, Xian and Xiaofei Lu (2013). A longitudinal study of receptive vocabulary breadth knowledge growth and vocabulary fluency development. *Applied Linguistics* 1-23. <<http://applied.oxfordjournals.org/content/early/2013/07/16/applin.amt014.full.pdf+html>> (Accessed November 6 2015).

APPENDIX I
VOCABULARY LEVELS TEST 2,000 (Schmitt, Schmitt and Clapham 2001)

2,000 WORD LEVEL TEST

CENTRO _____
CURSO _____ **GRUPO** _____ **FECHA** _____
APELLIDOS _____ **NOMBRE** _____

Este es un test de vocabulario. En la parte izquierda te presentamos grupos de seis palabras inglesas y a su derecha, los significados de sólo tres de ellas. **Escribe** junto a éstos, el **número** de la palabra inglesa correspondiente a dichos significados. Observa el siguiente ejemplo:

EJEMPLO

1 business
 2 clock _____ part of a house
 3 horse _____ animal with 4 legs
 4 pencil _____ something used for writing
 5 shoe
 6 wall



RESPUESTA CORRECTA

1 business
 2 clock 6 part of a house
 3 horse 3 animal with 4 legs
 4 pencil 4 something used for writing
 5 shoe
 6 wall

1 coffee
 2 disease _____ money for work
 3 justice _____ a piece of clothing
 4 skirt _____ using the law in the right way
 5 stage
 6 wage

1 adopt
 2 climb _____ go up
 3 examine _____ look at closely
 4 pour _____ be on every side
 5 satisfy
 6 surround

1 choice
 2 crop _____ heat
 3 flesh _____ meat
 4 salary _____ money paid regularly for doing a job
 5 secret
 6 temperature

1 bake
 2 connect _____ join together
 3 inquire _____ walk without purpose
 4 limit _____ keep within a certain size
 5 recognize
 6 wander

1 cap
 2 education _____ teaching and learning
 3 journey _____ numbers to measure with
 4 parent _____ going to a far place
 5 scale
 6 trick

1 burst
 2 concern _____ break open
 3 deliver _____ make better
 4 fold _____ take something to someone
 5 improve
 6 urge

1 attack
 2 charm _____ gold and silver
 3 lack _____ pleasing quality
 4 pen _____ not having something
 5 shadow
 6 treasure

1 original
 2 private _____ first
 3 royal _____ not public
 4 slow _____ all added together
 5 sorry
 6 total

1 cream
 2 factory _____ part of milk
 3 nail _____ a lot of money
 4 pupil _____ person who is studying
 5 sacrifice
 6 wealth

1 ancient
 2 curious _____ not easy
 3 difficult _____ very old
 4 entire _____ related to God
 5 holy
 6 social