

THE BENEFITS OF CLIL INSTRUCTION IN SPANISH STUDENTS' PRODUCTIVE VOCABULARY KNOWLEDGE

Andrés Canga Alonso
Mario Arribas García
Universidad de la Rioja

Resumen

Este artículo pretende (i) analizar el vocabulario productivo de 101 estudiantes españoles de inglés como lengua extranjera de 4º de ESO en dos centros de educación secundaria situados en el norte de España en dos tipos de instrucción AICLE y no-AICLE y (ii) comparar los resultados obtenidos por los participantes de acuerdo con su sexo. Para medir el vocabulario productivo de los estudiantes se utilizó la versión paralela del *Productive Vocabulary Levels Test (PVLТ)* (Laufer & Nation, 1995, 1999). Los resultados muestran que los alumnos AICLE obtienen puntuaciones significativamente mejores que sus compañeros no-AICLE siendo el tamaño del vocabulario productivo en ambos grupos inferior a 1000 palabras. En lo que respecta a las diferencias en cuanto al sexo no se constatan diferencias significativas entre los chicos y las chicas. Estos resultados parecen indicar que el enfoque AICLE resulta beneficioso para el aprendizaje del vocabulario productivo en una lengua extranjera.

Palabras clave: Educación Secundaria, Vocabulario productivo, AICLE, no-AICLE, sexo.

Abstract

This paper aims at (i) investigating the productive vocabulary knowledge of 101 10th grade (4th ESO) EFL Spanish students in two different types of instruction CLIL and non-CLIL, and (ii) analysing the results obtained according to sex-based differences among the participants. We used the parallel version of the Productive Vocabulary Levels Test (PVLТ) (Laufer & Nation, 1995, 1999) to measure students' productive vocabulary knowledge. Our results reveal that our CLIL sample obtained significantly better results than their non-CLIL partners, and the students' productive vocabulary size ranks below 1,000 words. As for sex-based differences, CLIL boys' mean scores are the highest, but the differences between both sexes, regardless of their type of instruction, are not statistically significant. These findings led us to believe that the CLIL approach offers a benefit for productive vocabulary learning in a foreign language.

Keywords: Secondary Education, productive vocabulary, CLIL, non-CLIL, sex differences.

1. Introduction

Vocabulary is acknowledged to be of paramount importance in foreign language learning (FLL). Hence, examining learners' word knowledge can provide interesting and reliable insights into their overall language knowledge. In recent decades, several studies have analysed the receptive vocabulary knowledge of EFL learners following an approach based on Content and Language Integrated Learning (CLIL), i.e. learning a content subject other than language lessons through the foreign language and traditional EFL (non-CLIL) instruction (Canga Alonso, 2013 a, b; Nikula, Dalton-Puffer & Llinares, 2013; Ruiz de Zarobe, 2011; Coyle, Hood & Marsh, 2010; Jiménez Catalán and Ruiz de Zarobe, 2009; Jiménez Catalán, Ruiz de Zarobe & Cenoz, 2006). However, to our

knowledge, there is a scarcity of research concerning productive vocabulary knowledge in CLIL and non-CLIL instruction at secondary school level in Spain. The main purpose of this paper is to compare general productive vocabulary size of traditional EFL and CLIL learners since it is generally believed that words are known receptively first and only after intentional or incidental learning become available for productive use. CLIL learners have a longer exposure to the foreign language, thus, the present research wants to prove that this exposure fosters productive vocabulary learning and CLIL subjects obtain better results in the Productive Vocabulary Levels Test (PVLVT) (Laufer & Nation, 1995, 1999).

The study also explores differences in productive vocabulary knowledge of male and female students. Recent studies on sex-based differences and vocabulary learning have not reached an agreement to assert that one sex outperforms the other as far as receptive and productive word knowledge is concerned. Thus, our findings will aim to shed some light on this aspect by comparing the scores obtained by male and female students in response to the PVLVT. Hence, we review the main studies dealing with the importance of vocabulary knowledge and sex-based differences and explore the relationship between CLIL and foreign language vocabulary. A report of the study conducted with its methodology, main results found and interpretation of those results follows. The paper concludes by pointing out some lines for further research trying to overcome the main limitations of the present study.

2. Vocabulary Knowledge and sex variables in foreign language learning

A large vocabulary size is essential to interacting in a foreign language. In this sense, researchers have tackled the issue concerning the number of words necessary to understand spoken discourse (Adolphs & Schmitt, 2004; Nation, 2001) and to read and comprehend texts in the native and foreign language (Laufer, 1997; Anderson & Freebody, 1981). Among the former researchers, Adolphs & Schmitt (2004) estimated that, at least, 2,000 word forms have to be mastered in order to understand around 90% and 94% of spoken discourse in different contexts. Among the latter, Laufer (1992, 1997) stated that a text coverage of 95% can be reached with a 5,000-word English vocabulary or 3,000 word families, which agrees with the assertions made by Hazenberg & Hulstijn (1996), Nation (1993, 2001) and Cobb & Horst (2004). More recently, Nation (2006) asserted that 8,000 to 9,000 word families are needed for understanding a written text and a vocabulary of 6,000 to 7,000 word families for comprehension of spoken texts, if 98% coverage of a text is desired. Hirsh & Nation (1992) also pointed out that knowledge of 5,000 word families is necessary to enjoy reading. Estimates based on word frequency criteria have been calculated and research claims that gaining command of the 2,000-3,000 most frequent words as soon as possible is vital for the language learner to communicate orally and in written form in the foreign language (Milton, 2009; Nation & Waring, 1997; Nation, 1993). The sooner the most frequent words are learned by students, the better their language performance will be. As Schmitt claims: “The learning of these basic words cannot be left to chance, but should be taught as quickly as possible, because they open [...] the door of further learning” (2000: 137).

As it has been mentioned, vocabulary size is one of the central dimensions in explorations of lexical proficiency. Different studies have set to the task of finding estimations of productive and

receptive vocabulary size. The tests designed for this purpose are generally built upon frequency lists (Nation, 1990, 2001) on the assumption that knowledge of less frequent words implies knowledge of more frequent words (Schmitt, 2000). Based on this idea of finding estimations of productive word knowledge, several instruments have been designed to measure productive vocabulary size. One of the most frequently used tests is Lex30 (Meara & Fitzpatrick, 2000). It is a free word association task which assesses discrete, comprehensive and context-independent vocabulary. It has several practical advantages since it generates a rich vocabulary output very economically, that is, through single word prompts. It is easily administered and it requires very little time to complete (15 minutes); and it is scored automatically using a computer programme (Jiménez Catalán & Moreno Espinosa, 2005). This test has been implemented in primary education (Moreno Espinosa, 2009; Jiménez Catalán & Moreno Espinosa, 2005) and with undergraduates (Fitzpatrick & Meara, 2004, Jiménez Catalán & Moreno Espinosa, 2004). However, it presents some methodological problems when classifying the words that the Lex30 scorer did not recognize in their appropriate bands (Jiménez Catalán & Moreno Espinosa 2005).

Another instrument which has been used widely to explore students' productive vocabulary knowledge is the Productive Vocabulary Levels Test (PVLТ) (Laufer & Nation, 1995, 1999). The PVLТ addresses two dimensions of lexical competency: partial productive word knowledge and word frequency. It measures discrete, selective and context dependent vocabulary (Mochizuki, 2012; Moreno Espinosa, 2010). One of the advantages of the test is that frequency bands are independent from one another, which allows researchers to give their testees either the whole test or only the bands that are suitable for their learners' level. This fact together with its format (see Appendix I) seems to be appropriate for secondary school students. Nevertheless, this test has not been as widely implemented as its receptive version (Schmitt, Schmitt & Clapham 2001), which has been used to test primary (Agustín Llach & Terrazas Gallego, 2012; Terrazas Gallego & Agustín Llach, 2009; Jiménez Catalán & Terrazas Gallego, 2005-2008) and secondary school (Canga Alonso, 2013a) Spanish students' receptive vocabulary knowledge in traditional EFL instruction.

Considering these facts, it was decided that the PVLТ is a good test to explore the productive vocabulary size of the sample of Spanish EFL learners in the last year of secondary school for two main reasons: it measures discrete, context dependent vocabulary, and it allows for profiling learners' vocabulary size on the basis of the frequency level to which the words used belong. Hence, the main aim of the present paper is to explore 10th grade Spanish students' productive vocabulary size in CLIL and non-CLIL instructional contexts, using the PVLТ as the instrument of measurement.

Having analysed the importance of vocabulary in L2 learning as well as having referred to the research conducted on vocabulary learning, the importance of sex-based differences in the literature on vocabulary learning can be considered. The role of sex has also occupied an outstanding place in current research on vocabulary in a FL. Receptive and productive vocabulary knowledge of male and female learners has been widely examined, and scholars have reached different conclusions. Boyle (1987) concludes that, exception to what, boys are superior to girls in the comprehension of heard vocabulary. Similarly, Scarcella & Zimmerman (1998) found that men performed significantly better than women in a test of academic vocabulary recognition, understanding and use. In Lynn, Fergusson, & Horwood (2005), and Edelenbos & Vinjé (2000), males also outperformed females in vocabulary

knowledge in the foreign language. By contrast, in Nyikos' study (1990) women performed better than men in a memorization test of German vocabulary. Nevertheless, Jiménez Catalán & Terrazas Gallego (2005-2008) discovered no significant sex-based differences in performance on a receptive vocabulary test implemented with primary students.

In the same vein, in a recent longitudinal research Agustín Llach & Terrazas Gallego (2012) found very slight differences among males and females across grades in the context of Spanish primary education concerning their receptive vocabulary knowledge. Canga Alonso (2013b) also found slight differences in his study of primary and secondary school students' receptive vocabulary. On the contrary, highly significant differences were found in favour of females in the mean number of words produced in response to the 15 cues of a lexical availability test (Jiménez Catalán & Ojeda Alba, 2009). A set of recent studies compiled in Jimenez Catalán (2010) also pointed to mixed results on sex differences or tendencies. In this same volume, García Gómez (2010) explored the development of gender relations and identities through analysis of writing samples from Spanish male and female EFL university students, and his findings revealed male and female gender-biased subjacent ideologies. As Sunderland (2010) claims, a careful analysis of this compilation provides the conclusion that the relationships between vocabulary and gender are not enduring, but may be context and test type-specific. These relationships can also be influenced by L1, age or L2 proficiency since the studies 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.

Considering the aforementioned studies, it can be stated that results are inconclusive regarding the role of sex in foreign language learning and in particular in lexical acquisition. Furthermore, the type of word knowledge explored, the learning context, or the task used for data gathering seem to play a relevant role in the establishment of sex tendencies. For this reason, the present study tries to ascertain if the scores obtained by male and female students in the PVLТ are alike since no difference has been made in the formal instruction they received either with a CLIL or a non-CLIL approach as both high schools mix boys and girls in the same classroom

3. CLIL and vocabulary learning

As mentioned in the introduction, CLIL implies the teaching of a content subject through English. Thus, its essence 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 & Frigols, 2008: 11). CLIL also provides real and meaningful input for the learner in form of subject content and language for classroom management (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 encourage communication in the classroom. Feedback is also integrated into classroom discourse to encourage interaction among apprentices. Thus, the challenge in a 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.

Vocabulary size 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 FL increases, less frequent words are incorporated into the lexicon (Milton 2009; Vermeer 2001; Barrow Nakanishi, & Ishino, 1999). Therefore, it seems evident that a content-based approach provides more opportunities to learn, either explicitly or implicitly, target vocabulary in meaningful situations (Pérez-Vidal, 2009; Muñoz, 2007), since learners are exposed to the target language for a longer period than students' enrolled in traditional EFL classrooms. Xanthou (2011) proved that CLIL had a positive impact in a group of primary school children in Cyprus regarding students' vocabulary tests results which demonstrates that by attaching words to their surroundings, the likelihood of comprehension and retention is increased. These gains in receptive vocabulary size align with other research conducted in Spain (Jiménez Catalán & Ruiz de Zarobe, 2009; Jiménez Catalán, Ruiz de Zarobe & Cenoz, 2006), where significant results were obtained in favour of the CLIL group in receptive vocabulary knowledge. In a similar study, Canga Alonso (2013a) found statistically significant differences between 6th grade primary students in CLIL contexts and those enrolled in a traditional EFL approach. Nevertheless, to our knowledge, there is a lack of research in productive vocabulary knowledge in CLIL and non-CLIL types of instruction using the PVL in Spanish secondary school. For these reasons and the aspects mentioned in the previous section, this study aims at (i) investigating the productive vocabulary knowledge of 15-16 year-old male and female Spanish students learning English in Spanish 10th grade (4th ESO) in relation to: ii) type of instruction (CLIL vs. non-CLIL) and iii) sex-based differences. Thus, the researchers set out to find answers to the following research questions:

1. What is the productive vocabulary knowledge of 10th grade EFL learners in CLIL and non-CLIL instruction?
2. Are there any statistically significant differences in productive vocabulary sizes between CLIL and non-CLIL learners?
3. Do we find statistically significant differences in students' performance with regard to sex in the Productive Vocabulary Levels Test?

4. Method

4.1 Participants

Two student samples constitute the participants of this study. The CLIL group, is made up of 73 learners (37 boys and 36 girls), whereas the traditional or non-CLIL group, comprises 38 learners (26 boys and 12 girls). Both groups learn English as a curricular subject in the same grade of secondary education. (4th ESO/10th Grade). They have been learning English as a school subject since the 1st year of primary education in two weekly sessions of 50 to 60 minutes. Additionally, the CLIL group has been receiving extra-exposure to English since their 7th grade (1st year of Secondary Education).

The sample is homogeneous in regards to their socio-economic and cultural background, since participants were drawn from two high schools in the same area. Students also shared Spanish as their

native language (L1). The groups differ in the kind of instruction they received, i.e. CLIL vs. non-CLIL, and consequently, in the number of hours of exposure to English FL. Learners in the non-CLIL group were exposed to English through the English FL school subject, exclusively. However, learners in the CLIL group received, apart from the weekly EFL lessons, input in Natural Sciences and Arts and Crafts through the medium of English. Consequently, not only the amount but also the nature of the input differs between the traditional and the CLIL group. Traditional learners have received approximately 1,049 hours of exposure to EFL on a yearly basis since 1st grade of primary. The CLIL group has received the same 1,049 hours plus 30-60 more hours in CLIL.

Table 1 illustrates the approximate number of hours of exposure students have received by the time of data collection.

		Hours of exposure	
Grade	Age	CLIL	Non- CLIL
4 th Secondary	15-16	1,079-1,109	1,049

Table 1. Hours of exposure to English FL

4.2 Data Collection

The 2,000 word parallel version (version A+ version C) of the Productive Vocabulary Levels Test (PVLТ) (see Appendix 1) was used to measure the productive vocabulary knowledge of these subjects (Laufer & Nation 1999; Laufer & Nation 1995). The test measures knowledge of vocabulary at the 2,000, 3,000, 5,000 and 10,000 word bands (e.g. the 2000 word band would test words 1 – 2,000). The word bands correspond to the 2,000 and so on most frequent words in English which are based on West's (1953) General Service List and the Thorndike & Lorge's list (1944). We have chosen the 2,000 band of the test since as mentioned in section one, gaining command of the 2,000 most frequent words as soon as possible is vital for the language learner to communicate orally and in written form in the foreign language (Milton, 2009; Nation, 1993; Nation & Waring, 1997). The PVLТ measures controlled productive knowledge (Laufer, 1998) since test-takers have to complete a missing word in 30 different sentence contexts where they are given the first letters of the target word as a cue. Therefore, the PVLТ is a reliable, valid and practical measure of vocabulary growth (Laufer & Nation, 1999: 44). Completing the test requires the knowledge of meaning, form, phonological aspect and collocations of the target word. In addition, reading comprehension is required in gaining the clues to complete the task.

Data was collected in one session during school time. 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' L1 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 (the relevant word list divided by N items in test)" (Nation, 1990: 78) was applied. In order for an answer to be

correct, the word has to be both grammatically and orthographically adequate, e.g. if the missing word is a verb, it has to be written in the corresponding tense. Therefore, a wrong verb tense would get 0 points in that given sentence.

The sample was also analysed with SPSS 19 to check whether there were statistically significant differences according to type of instruction, sex and productive vocabulary knowledge.

5. Results

Regarding our first research question (What is the productive vocabulary knowledge of 10th grade EFL learners in CLIL and non-CLIL instruction?) the data (see table 2) revealed that the maximum score in the sample was twenty-nine points out of thirty which was attained by one of the students in the CLIL group whereas the minimum score (zero out of thirty) was achieved by one participant in the non-CLIL group. These figures indicate that students are behind from learning the 2,000 most frequent words according to the frequency lists collected by Kucera & Francis (1967), West (1953), and Thorndike & Lorge (1944). The data also implies that students can have problems understanding spoken and written discourse in English since they need to have gained at the least the 2,000 most frequent words in order to communicate orally and in written form in the foreign language (Nation & Waring, 1997). Mean scores are also higher in favour of the CLIL group (12.19 vs. 9.6) but standard deviations are quite similar for both cohorts (4.9 vs.5). This implies that the variability of results is quite alike for both groups.

	Min.	Max.	Mean	SD
CLIL (n=73)	3	29	12.19	4.9
Non-CLIL (n=38)	0	22	9.6	5

Table 2. PVLTL 2,000 results.

This profile is illustrated in the rankings of percentages summarized in figure 1. The results show that our CLIL informants obtained better results than our non-CLIL students in all the ranks but the lowest (0-5). It is also noteworthy that both groups of students attained similar scores in the 6-10 and 11-15 ranks, although CLIL students performed slightly better. These two ranks (6-10 and 11-15) are also the ones with the highest percentages for both cohorts. On the contrary, the ranking 26-30 is quite low in the CLIL group and shows no evidence in the non-CLILs.

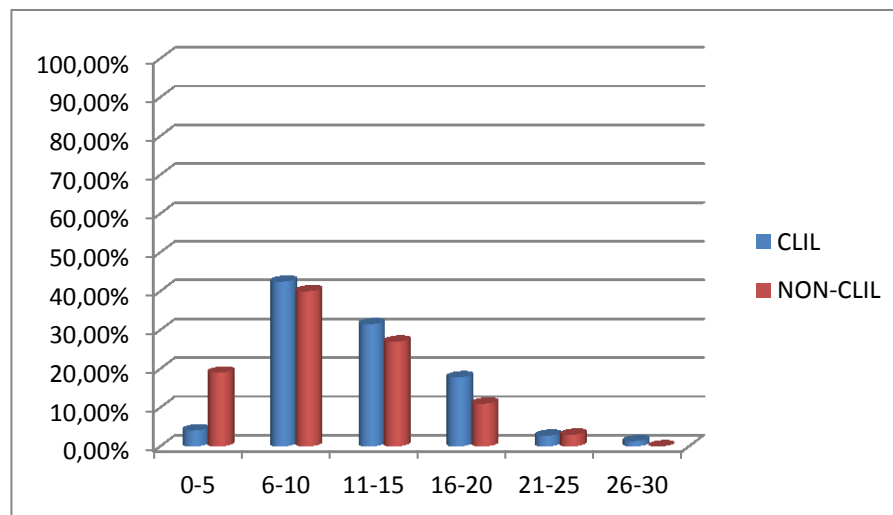


Figure 1. Frequency distribution of tests scores CLIL and non-CLIL groups.

As for the second research question (Are there any statistically significant differences in productive vocabulary sizes between CLIL and non-CLIL learners?), students' scores were translated into the number of known words for each frequency level applying Nation's formula (1990: 78). As illustrated in table 3, the means obtained by the CLIL group confirms our previous presupposition on a two-fold basis: CLIL learners recognized and produced a higher number of words (813) when compared to the non-CLIL group (640) and both groups were under the 1,000 most frequent words in English. The standard deviations are quite alike, although the CLIL cohort presents greater variability in their scores and their deviations are slightly higher when compared to the non-CLIL cohort (327 vs 304)

	Min.	Max.	Mean	SD
CLIL (n=73)	200	1933	813	327
Non-CLIL (n=38)	0	1467	640	304

Table 3. Estimation of words CLIL and non-CLIL groups.

Kolmogorov-Smirnov and Shapiro Wilk tests were implemented in order to ascertain whether our sample met the normality assumption. The values obtained (see table 4) indicate their distribution is not normal for the variable type of instruction, therefore non-parametric tests had to be implemented.

Test	Type of instruction	Kolmogorov-Smirnov			Shapiro-Wilk		
		Statistics	gl	Sig.	Statistics	gl	Sig.
PVL	CLIL	.138	73	.002	.950	73	.006
	Non-CLIL	.133	38	.087	.969	38	.368

Table 4. Parametric tests: Type of Instruction

The Mann-Whitney U test was applied in order to ascertain our results were statistically significant regarding the two types of instruction analysed in this paper. As illustrated in table 5 The

p-value ($p=0.014$) is very close to 0.01, which proves that there are statistically significant differences between CLIL and non-CLIL learners' word estimates in favour of the CLIL group.

	PVLT
Mann-Whitney U	994.50
Wilcoxon W	1735.50
Z	-2.445
P. (two tailed)	.014

Table 5. Non-parametric tests: Type of instruction

The results for our third research question (Do we find statistically significant differences in students' performance with regard to sex and type of instruction in the Productive Vocabulary Levels Test?), CLIL girls obtained better maximum scores (29 points) in the PVLT whereas non-CLIL boys attained the highest rate in the lowest scale of the rank (i.e. 0-5 points). As can be seen in Figure 2, most of the informants, regardless of their sex and type of instruction scored between 6-10 points in the productive vocabulary test. With regard to medium scale of the rank (i.e. 11- 15 points), non-CLIL and CLIL girls obtained the highest percentages (33%). On the contrary, CLIL boys are much better in regards to the 16-20 point rank (24% vs. 11%). Finally, 5% of non-CLIL and CLIL boys scored more than 20 points but less than 25 in the PVLT. As we will analyse in the following section, these findings show the tendency that boys and girls behave similar to peers of their same sex regardless of the type of instruction.

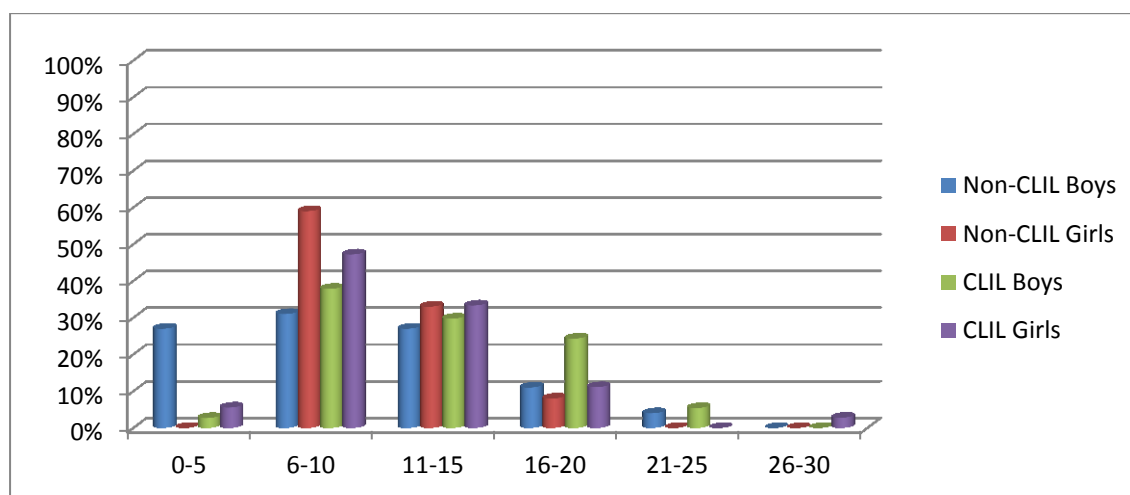


Figure 2. Frequency distribution of tests scores for boys and girls.

The box-plot in figure 3 illustrates graphically the median values for boys and girls in both types of instruction, which concord with the data analysed above. This figure also indicates that our results are quite homogeneous for both cohorts. There is also an outlier which belongs to a girl whose scores are much higher than the ones obtained by the rest of the students. She belongs to the CLIL group and is the informant who scored 29 points in the PVLT:

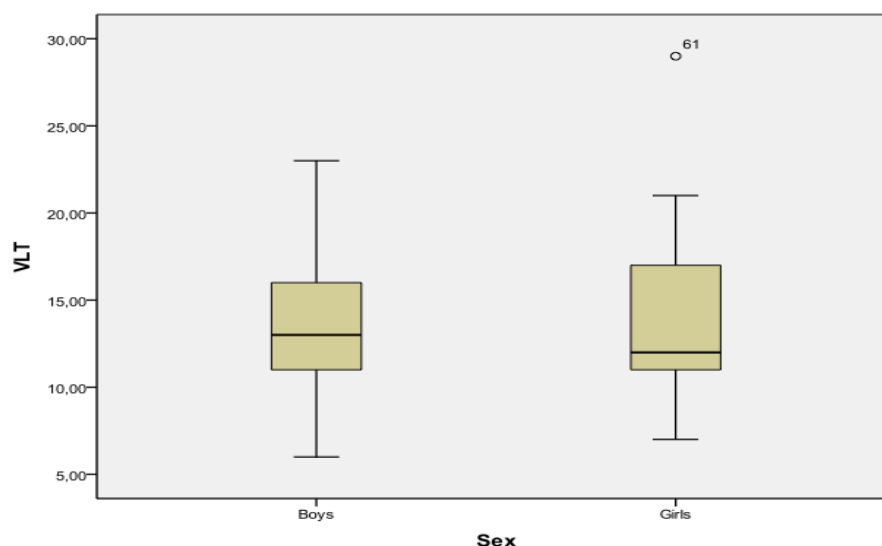


Figure 3. Median vales according to sex and type of instruction.

Regarding word estimates and sex-based differences, the descriptive statistics shown in table 6 illustrate that non-CLIL boys obtained the lowest mean scores (635 words) closely followed by non-CLIL girls (661 words). One CLIL girl obtained the highest score (1933 words), but CLIL boys attained the highest mean score (876 words).

		Min	Max	Mean	SD
CLIL	Boys (n=37)	267	1400	876	301
	Girls(n=36)	200	1933	748	344
Non-CLIL	Boys (n=26)	0	1467	635	343
	Girls(n=12)	400	1067	661	240

Table 6. Distribution of word estimates according to sex.

Kolmogorov-Smirnov and Shapiro-Wilk parametric tests were implemented in order to ascertain whether our sample met the normality assumption according to sex based differences. As shown in table 7, the p-values obtained were lower than ($p=0.05$) for the girls, so the sample did not meet normality and, therefore, non-parametric tests were applied.

Test	Sex	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistics	gl	Sig.	Statistics	gl	Sig.
PV LT	Boy	.100	63	.193	.981	63	.432
	Girl	.131	48	.039	.943	48	.020

Table 7. Parametric tests for sex-based differences.

The U Mann-Whitney test was conducted to calculate inferential statistical differences among our informants according to sex. Its results reveal that there are not significant differences at a significance level of 5% ($p=0.13$) in vocabulary size estimations. Table 8 offers these results:

	PVLT
Mann-Whitney U	1260.50
Wilcoxon W	3276.50
Z	-1.50
P. (two tailed)	.134

Table 8. Results of inferential statistics for sex-based differences.

6. Discussion

The analysis of our data clearly shows that, regardless of the type of instruction (i.e. CLIL/non-CLIL), the productive vocabulary knowledge of the 10th graders analysed in the present research is lower than 1,000 words. If we compare this data with the findings obtained by Moreno Espinosa (2010) in the same educational level and socio-cultural context, we ascertain that productive vocabulary size is practically the same for both non-CLIL samples of students since our students' estimation of words is 640 and Moreno Espinosa's 645. These results may also imply that the EFL instruction Moreno Espinosa's and our sample of non-CLIL groups of learners have received in the compulsory schooling, is rather similar, regardless of the type of school students have attended.

Our data also indicates that CLIL learners obtained significantly higher scores than non-CLIL students in the PVLT ($p=0.014$), which seems to show that CLIL instruction may favour productive vocabulary learning. This result is in line with previous studies, which show that CLIL or longer foreign language exposure programs foster vocabulary learning, and that benefits start appearing after some time (Agustín Llach and Canga Alonso, 2014; Celaya & Ruiz de Zarobe, 2010). The longer exposure to English input CLIL learners have received can help explain this advantage in general productive 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 an avenue for further research, future studies could test learners productive knowledge of specific vocabulary related to certain fields such as Natural Sciences and/or Arts and Crafts (the CLIL subjects). We could speculate that results would favour the CLIL group, whose productive vocabulary size might be higher for both semantic fields. However, this is just mere speculation, since we have not conducted such test and further research on words students may have acquired in the content classes together with a comparison of their results to non-CLIL learners' is needed.

In short, CLIL seems to favour productive vocabulary knowledge, but there might be other factors influencing students' productive vocabulary learning, such as exposure time. It is indeed difficult to discern whether the CLIL factor or the longer exposure (more instruction hours) is responsible for the CLIL advantage in our data, as these two factors are impossible to untangle in our present study.

If we compare our findings with previous studies on receptive vocabulary size conducted in the same area and educational level but with a different sample of students Canga Alonso (2013c), we ascertain that the estimations of words are higher in the receptive vocabulary test (935 words). These results concord with previous research on students' receptive and productive vocabulary sizes as students tend to obtain better results in receptive tests than in productive ones because reception is a prior step to production (Zhong, 2012; Yamamoto 2011; Martínez Adrián & Gallardo del Puerto, 2010; Webb 2008; Laufer, 1998). However, this comparison should be taken with caution since Canga Alonso has not conducted research on productive vocabulary knowledge and his sample of students is different from the one analysed in the present study. As abovementioned, research has shown that receptive skills are developed prior to productive ones which explains that students' results in the receptive VLT were higher. Therefore, further research is called to correlate the receptive and productive vocabulary of the students in the present sample to compare it with previous studies on receptive and productive vocabulary in the same area.

Our results also indicate that CLIL and non-CLIL learners would have problems in performing certain linguistic tasks such as understanding informal spoken discourse for which Adolphs & 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. Nevertheless, EFL learners would have an easier time watching TV programmes (Webb & Rogers 2009a), or movies (Webb & Rogers 2009b), since the first 1,000 most frequent words make up for around 85% of the total word coverage. Accordingly, researchers call for the command of the 2,000-3,000 most frequent words as soon as possible (Webb and Chang 2012; Nation & Waring, 1997; Nation, 1993), and our students' mean scores (813 vs 640 words) indicates that they have not even acquired the 1,000 most frequent words in English. Furthermore, we agree with Schmitt (2000) and Webb & 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 ($p=0.13$) although CLIL boys obtained the highest scores (876 words). This finding is in line with the results obtained in receptive vocabulary with students of their same age in non-CLIL instruction (Canga Alonso, 2013c), but differs from tests with younger CLIL and non-CLIL learners in their same educational context (Canga Alonso, 2013a, 2013b; Agustín Llach and Terrazas Gallego, 2012; Jiménez Catalán and Terrazas Gallego, 2005-2008) where girls obtained the highest scores. It is also noteworthy that the sample of CLIL students shows the opposite tendency since non-CLIL girls slightly outperformed non-CLIL boys (661 vs. 635 words) and followed CLIL girls (748 vs. 661 words). In light of these results on sex-based differences and type of instruction, we agree with Sunderland's (2010) assertion that the relationships between vocabulary and sex are not enduring, but may be context and test type-specific, and also influenced by L2 proficiency since by implementing the same test (i.e. PVLТ parallel version A+C) in the same context (i.e. last year of Spanish compulsory education) students obtained different results according to their sex and type of instruction. In this way, male CLIL learners obtained the highest scores whereas non-CLIL girls outperformed their male non-CLIL partners. These similar results for both sexes also verify our initial presupposition that sex-based scores should be similar since both sexes were treated equally in the classroom and they received the same formal instruction.

All in all, this data indicates that CLIL instruction has been beneficial for the CLIL group. However, we should take this assumption with caution since the sample of CLIL learners is not very numerous and further research is needed in order to test if a bigger sample of CLIL students would obtain similar results to the ones shown in this paper.

7. Conclusions

Three main findings stand out from the present research study. First, the productive vocabulary size of our sample of Spanish secondary school CLIL and non-CLIL learners lies below the range of first 1,000 most frequent words in English, according to the results of the PVLIT, which shows that our students might find it difficult to understand spoken and written texts in English. As expected, their scores are lower to those obtained by students of their same age and sociocultural background in receptive vocabulary tests since reception is a previous step to production when you learn a foreign language. With regard to the second research question, the results purport that CLIL instruction seems to favour larger productive vocabulary sizes as CLIL learners obtained significantly better results than non-CLILs. Finally, no statistically significant differences can be established according to the sex of the participants, which implies that boys and girls have added a similar amount of words to their productive vocabulary. Nevertheless, these results should be taken with caution due to the number of students who took part in the present research. Therefore, further research is needed in order to explore if the productive vocabulary size of a bigger sample of students of the same age and types of instruction from different schools in the area remains alike or is significantly higher or lower than our informants' productive vocabulary knowledge. One further limitation of the present study is the use of a single, and somewhat limited, instrument to measure productive vocabulary size. Using other tests for vocabulary knowledge, such as lexical availability tests might provide even more insightful results and reveal more qualitative data concerning learners' vocabulary knowledge. Finally, CLIL programmes have spread in Spain in the last decade; therefore it would be interesting to relate 10th graders receptive and productive vocabulary knowledge in order to ascertain which type of instruction (CLIL/non-CLIL) could benefit vocabulary learning. Further research is called for to overcome these limitations and address this new research path on CLIL instruction.

Acknowledgements

We would like to acknowledge the financial support of the Spanish Ministry of Science and Innovation through grant nº FFI 2010-19334.

References

- Adolphs, S. & Schmitt, N. 2004. "Vocabulary coverage according to spoken discourse context". In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language* (pp. 39-52). Amsterdam: John Benjamins.
- Agustín Llach, M^a P. & Canga Alonso, A. 2014. "Vocabulary growth in young CLIL and traditional EFL learners: evidence from research and implications for education". *International Journal of Applied Linguistics*. DOI: 10.1111/ijal.12090.
- Agustín Llach, M^a P. & Terrazas Gallego, M. 2012. "Vocabulary knowledge development and gender differences in a second language". *ELIA*, 12, 45-75.
- Anderson, R. C. & Freebody, P. 1981. "Vocabulary knowledge. In J. Guthrie (Ed.), *Comprehension and teaching: Research reviews* (pp. 203-226). Newark, DE: International Reading Association.

- Barrow, J., Nakanishi, Y. & Ishino, H. 1999. "Assessing Japanese college students' vocabulary knowledge with a self-checking familiarity survey". *System*, 27, 223-247.
- Boyle, J. P. 1987. "Sex differences in listening vocabulary". *Language Learning*, 37(2), 273-284.
- Canga Alonso, A. 2013a. "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. Retrieved from: doi: 10.5294/laclil.2013.6.2.2.
- Canga Alonso, A. 2013b. "The receptive vocabulary of Spanish students in CLIL and non-CLIL instruction". *Op. cit.*: Series II (2). Available from <https://docs.google.com/viewer?a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbXhcGVhYWRpcmVjYW98Z3g6MWZmMGEwODAwMmMyMGRiYg>
- Canga Alonso, A. 2013c. "Receptive vocabulary size of Spanish EFL learners at secondary school." *Revista de Lingüística y Lenguas Aplicadas*, 8, 66-75. Retrieved from: doi: <http://dx.doi.org/10.4995/rlyla.2013.1180>
- Celaya, M^a L., & Ruiz de Zarobe, Y. 2010. "First languages and age in CLIL and non-CLIL contexts". *International CLIL Research Journal*, 1(3). Retrieved from <http://www.icrj.eu/13/article6.htm>.
- Cobb, T., & Horst, M. 2004. "Is there room for an academic word list in French?" In P. Bogaards & B. Laufer (Eds.), *Vocabulary in a second language* (pp. 15-38). Amsterdam: John Benjamins.
- Coyle, D., Hood, P., & Marsh, D. 2010. *CLIL: Content and Language Integrated Learning*. Cambridge: Cambridge University Press.
- Dalton-Puffer, C. 2007. *Discourse in Content and Language Integrated Learning (CLIL)*. Amsterdam/Philadelphia: John Benjamins.
- Dalton-Puffer, C. 2008. "Outcomes and processes in Content and Language Integrated Learning (CLIL): Current research from Europe". In Delanoy, W., & Volkman, L (Eds.), *Future perspectives for English language teaching* (pp. 139-157). Heidelberg: Carl Winter.
- Edelenbos, P., & Vinjé, M. 2000. "The assessment of a foreign language at the end of primary (elementary) education". *Language Testing*, 17(2), 144-162.
- Fitzpatrick, T., & Meara, P. 2004. "Exploring the validity of a test of productive vocabulary". *Vigo International Journal of Applied Linguistics (VIAL)*, 1, 55-73. Retrieved from <http://webs.uvigo.es/vialjournal/pdf/Vial-2004-Article3.pdf>
- García Gómez, A. 2010. "Lexical encoding of gender relations and identities". In R. M. Jiménez Catalán (Ed.), *Gender Perspectives on Vocabulary in Foreign and Second Languages* (pp. 238-63). Basingstoke: Palgrave Macmillan.
- Hazenber, S., & Hulstijn, J. H. 1996. "Defining a minimal receptive second-language vocabulary for non-native university students: An empirical investigation". *Applied Linguistics*, 17, 145-63.
- Hirsh, D., & Nation, P. S. N. 1992. "What vocabulary size is needed to read unsimplified texts for pleasure?" *Reading in a Foreign Language*, 8, 689-696.
- Jiménez Catalán, R. M^a. (Ed.). 2010. *Gender perspectives on vocabulary in foreign and second language*. Basingstoke: Palgrave Macmillan.
- Jiménez Catalán, R. M^a. & Moreno Espinosa, S. 2004." L2 word associations and the variable sex: An outline according to and electronic tool". In A. R. Celada, D. Pastor García and P. J. Pardo García, (Eds.), *Proceedings of the 27th International AEDEAN Conference*. (pp. 54-77). Salamanca: Ambos Mundos.
- Jiménez Catalán, R. M^a. & Moreno Espinosa, S. 2005. "Using Lex30 to measure the L2 productive vocabulary of Spanish primary learners of EFL". *Vigo International Journal of Applied Linguistics (VIAL)*, 2, 27-44. Retrieved from <http://webs.uvigo.es/vialjournal/pdf/Vial-2005-Article2.pdf>
- Jiménez Catalán, R. M^a. & Ojeda Alba, J. 2009. "Girls' and boys' lexical availability in EFL". *International Journal of Applied Linguistics*, 158, 57-76.
- Jiménez Catalán, R. M^a, & Ruiz de Zarobe, Y. 2009. "The receptive vocabulary of EFL learners in two instructional contexts: CLIL versus non-CLIL". In R. M^a Jiménez Catalán & Ruiz de Zarobe, Y. (Eds.), *Content and language integrated learning. Evidence from research in Europe* (pp. 81-92). Bristol: Multilingual Matters.
- Jiménez Catalán, R. M^a, Ruiz de Zarobe Y., & Cenoz, J. 2006. "Vocabulary profiles in English as a subject and as a vehicular language". *VIEWZ*, 15 (3), 23-27. Retrieved from http://anglistik.univie.ac.at/fileadmin/user_upload/dep_anglist/weitere_Uploads/Views/views15_3_clil_special.pdf.

- Jiménez Catalán, R. M^a. & Terrazas Gallego, M. 2005-2008. "The receptive vocabulary of English foreign language young learners". *Journal of English Studies*, 5, 173-191.
- Kucera, H. & Francis, W. N. 1967. *A computational analysis of present day American English*. Providence: Brown University Press.
- Laufer, B. 1992. "How much lexis is necessary for reading comprehension". In P. J. L. Arnaud & H. Béjoint (Eds.), *Vocabulary and applied linguistics* (pp. 126-132). London: Macmillan.
- Laufer, B. 1997. "The lexical plight in second language reading: Words you don't know, words you think you know, and words you can't guess". In J. Coady & T. Huckin (Eds.), *Second language vocabulary acquisition* (pp.20-34). Cambridge: Cambridge University Press.
- Laufer, B. 1998. "The development of passive and active vocabulary in a second language: Same or different?" *Applied Linguistics*, 19(2), 255-271.
- Laufer, B., & Nation, P. S. N. 1995. "Vocabulary size and use: lexical richness in L2 written production". *Applied Linguistics*, 16, 307-322.
- Laufer, B., & Nation, P. S. N. 1999). "A vocabulary-size test of controlled productive vocabulary". *Language Testing*, 16(1), 33-51.
- Lynn, R., Fergusson, D. M. & Horwood, L. J. 2005. "Sex differences on the WISC-R in New Zealand". *Personality and Individual Differences*. 39, 103-114.
- Martínez Adrián, M. & Gallardo del Puerto, F. 2010." L2 receptive and productive vocabulary in senior learners". *ITL. International Journal of Applied Linguistics*, 160, 112-131.
- Mehisto, P., Marsh, D. & Frigols, M^a. J. 2008 *Uncovering CLIL: Content and Language Integrated Learning in bilingual and multilingual contexts*. Oxford: Macmillan.
- Meara, P., & Fitzpatrick, T. 2000. "Lex 30: An improved method of assessing productive vocabulary in an L2. *System*, 28, 19-30.
- Milton, J. 2009. *Measuring second language vocabulary acquisition*. Bristol: Multilingual Matters.
- Mochizuki, M. 2012. "Four empirical vocabulary tests studies in the three dimensional framework". *Vocabulary Learning and Instruction*, 1(1), 44-52. Retrieved from <http://www.vli-journal.org/issues/01.1/issue01.1.08.pdf>
- Moreno Espinosa, S. 2009. "Young learners' L2 word association responses in two different learning contexts". In R. M^a Jiménez Catalán & Y. Ruiz de Zarobe (Eds.), *Content and Language Integrated Learning in Europe* (pp. 241-256). Clevedon: MultilingualMatters.
- Moreno Espinosa, S. 2010. *Evaluación del vocabulario en redacciones escritas por aprendices de inglés como L2 a través de la tecnología*. Unpublished doctoral dissertation, University of La Rioja.
- Muñoz, C. 2007. "CLIL. Some thoughts on its psychological principles". *RESLA*, 11-17.
- Nation, P. S. N. 1990. *Teaching and learning vocabulary*. New York: Newbury.
- Nation, P. S. N. 1993. "Using dictionaries to estimate vocabulary size: Essential, but rarely followed, procedures". *Language Testing*, 10, 27-40.
- Nation, P. S. N. 2001. *Learning vocabulary in another language*. New York: Cambridge University Press.
- Nation, P. S. N. 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, P. S. N. & Waring, R. 1997. "Vocabulary size, text coverage and word lists". In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, acquisition and pedagogy* (pp. 6-19). New York: Cambridge University Press.
- Nikula, T, Dalton-Puffer, C, & Llinares, A. 2013. "CLIL classroom discourse. Research from Europe". *Journal of Immersion and Content- Based Language Education*, 1 (1), 70-100. Retrieved from doi: 10.1075/jib.1.1.04nik
- Nyikos, M. 1990. "Sex-related differences in adult language learning: Socialization and memory factors". *Modern Language Journal*, 74, 273-287.
- Pérez-Vidal, C. 2009. "The integration of content and language in the classroom: A European approach to education (the second time around)". In E. Dafouz, & M. C. Guerini (Eds.), *CLIL Across Educational Levels* (pp. 3-16). Madrid: Richmond.
- Ruiz de Zarobe, Y. 2011. "Which language competencies benefit from CLIL? An insight into applied linguistics research". In Y. Ruiz de Zarobe, J. M. Sierra and F. Gallardo del Puerto (Eds.), *Content and Foreign Language Integrated Learning*(pp. 123-153). Bern: Peter Lang.

- Scarcella, R. & Zimmerman, C. 1998. "Academic words and gender .ESL student performance on a test of academic lexicon". *Studies in Second Language Acquisition*, 20, 27-49.
- Schmitt, N. 2000. *Vocabulary in language teaching*. Cambridge: Cambridge University Press.
- Schmitt, N., Schmitt, D. & Clapham, C. 2001. "Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test". *Language Testing*, 18, 55-88. Retrieved from [http://www.norbertschmitt.co.uk/uploads/schmitt-n-schmitt-d-and-clapham-c-\(2001\)-developing-and-exploring-the-behaviour-of-two-new-versions-of-the-vocabulary-levels-test-language-testing-18-1-55-88.pdf](http://www.norbertschmitt.co.uk/uploads/schmitt-n-schmitt-d-and-clapham-c-(2001)-developing-and-exploring-the-behaviour-of-two-new-versions-of-the-vocabulary-levels-test-language-testing-18-1-55-88.pdf)
- Sunderland, J. 2010. "Theorizing gender perspectives in foreign and second language learning". In R. M. Jiménez Catalán (Ed.), *Gender Perspectives on Vocabulary in Foreign and Second Languages* (pp. 1-22). Basingstoke: Palgrave Macmillan.
- Terrazas Gallego, M. & Agustín Llach, M^a P. 2009. "Exploring the increase of receptive vocabulary knowledge in the foreign language: A longitudinal study". *IJES*, 9(1), 113-133. Retrieved from <http://revistas.um.es/ijes/article/viewFile/90681/87481>
- Thorndike, E. & Lorge, I. 1944. *The teacher's word book of 30,000 words*. New York Teachers College: Columbia University.
- Vermeer, A. 2001. "Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input." *Applied Psycholinguistics*, 22, 217-34.
- Webb, S. 2008. "Receptive and productive vocabulary sizes of L2 learners". *Studies in Second language acquisition*, 30(1), 79-95.
- Webb, S. & Rodgers, M. P. H. 2009a. "The vocabulary demands of television programs". *Language Learning*, 59(2), 335-366.
- Webb, S. & Rodgers, M. P. H. 2009b. "The lexical coverage of movies". *Applied Linguistics*, 30 (3), 407-427.
- Webb, S. & Chang, A. C-S. 2012. "Second language vocabulary growth". *RELC Journal*, 43(1), 113-126.
- West, M. 1953. *A general service list of English words*. London: Longman.
- Xanthou, M. 2011. "Current trends in L2 vocabulary learning and instruction. Is CLIL the right approach? " *Advances in research on language acquisition and teaching: Selected Papers* (pp. 459-471). Retrieved from <http://www.enl.auth.gr/gala/14th/Papers/English%20papers/Xanthou.pdf>
- Yamamoto, Y. 2011. "Bridging the gap between receptive and productive vocabulary size through extensive reading". *The Reading Matrix*, 11(3), 226-242.
- Zhong, H. 2012. "Multidimensional vocabulary knowledge: Development from receptive to productive use". In D. Hirsh (Ed.), *Current Perspectives in Second Language Vocabulary Research* (pp. 23-55). Bern: Peter Lang.

Andrés Canga Alonso holds a PhD in English Philology and is currently a lecturer at the University of La Rioja. His research focuses on the applications of e-mail tandem and autonomy in mixed-ability EFL learners, the analysis of receptive and productive vocabulary of primary and secondary CLIL and non-CLIL learners, and the use of language portfolios in EFL and translation. He has published several articles in scientific journals on the abovementioned fields, and is a member of GLAUR research group.

Mario Arribas García is an EFL teacher in a secondary school, a CLIL researcher, and teacher trainer. He is currently finishing his PhD thesis focused both on vocabulary teaching and learning in CLIL and EFL contexts, and the differences between CLIL and EFL textbooks and teaching materials. He has also published previous articles about vocabulary acquisition.

Appendix 1: Productive Vocabulary: Parallel Version (A+C) (Laufer & Nation, 1995 1999)

In the following sentences we have omitted the end of a word. Complete the sentences with the right word. For examples: *He was riding a bic_____*; you should have completed the sentence as follows: *He was riding a bicycle*.

The 2,000-word level

Example: He was riding a bicycle.

1. They will restore the house to its orig_____ state.
2. Each room has its own priv_____ bath and WC.
3. The tot_____ number of students at the university is 12,347.
4. They met to ele_____ a president.
5. Many companies were manufac_____ computers.
6. The lakes become ice-free and the snow mel_____.
7. They managed to steal and hi_____ some knives.
8. I asked the group to inv_____ her to the party.
9. She shouted at him for spoi_____ her lovely evening.
10. You must spend less until your deb_____ are paid.
11. His mother looked at him will love and pri_____.
12. The wind roa_____ through the forest.
13. There was fle_____ and blood everywhere.
14. She earns a high sal_____ as a lawyer.
15. The sick child had a very high tempe_____.
16. The bir_____ of her first child was a difficult time.
17. My favourite spo_____ is football.
18. In A.D. 636 an Arab army won a famous vic_____ over another army.
19. I'm glad we had this opp_____ to talk.
20. There are a doz_____ eggs in the basket.
21. Every working person must pay income t_____.
22. The pirates buried the trea_____ on a desert island.
23. Her beauty and ch_____ had a powerful effect on men.
24. La_____ of rain led to a shortage of water in the city.
25. He takes cr_____ and sugar in his coffee.
26. Pup_____ must hand in their papers by the end of the week.
27. Ann intro_____ her boyfriend to her mother.
28. Teenagers often adm_____ and worship pop singers.
29. In order to be accepted into the university, he had to impr_____ his grades.
30. The dress you're wearing is lo_____.