



Lexical Availability in CLIL: Differences in Language of Instruction

LEAH GEOGHEGAN*
University of La Rioja (Spain)

Received: 28/06/2023. Accepted: 31/01/2024.

ABSTRACT

Within the context of Content and Language Integrated Learning (CLIL), vocabulary acquisition is of central importance. However, while CLIL is increasingly being implemented throughout Spain, there remains a clear preference, both in practice and research, for using English over other languages such as French. This study thus investigates the token production of Spanish students taking both English and French CLIL classes by means of a lexical availability task. It aims to determine if there are quantitative differences between the learners' language level and token production in English and French, and whether these differences exist across grades. Results indicate that learners have higher proficiency and produce more words in English, and that CLIL instruction has a clear impact on learners' production across grades. The results are of key interest to multilingual CLIL educators seeking to make the most of vocabulary gains in multiple languages.

KEYWORDS: Content and Language Integrated Learning; Vocabulary; Lexical Availability; English as a Foreign Language; French as a Foreign Language.

1. INTRODUCTION

Since the turn of the century, Content and Language Integrated Learning (CLIL) has become increasingly widespread throughout Spain (Ruiz de Zarobe & Lasagabaster, 2010). The move towards this approach, whereby content classes are taught through a foreign language (FL), largely came about due to the general dissatisfaction with FL learning approaches in the 1980s and 1990s (Goris et al., 2019). This need for innovation and improvements in second language (L2) competencies led to the 1995 White Paper *Teaching and Learning: towards the Learning*

**Address for correspondence:* Departamento de Filologías Modernas, Universidad de La Rioja, c/ San Jose de Calasanz s/n, 26004 Logroño, La Rioja; email: legeogw@unirioja.es

Society (Eurydice, 2006), which stated that all individuals should be able to communicate not only in their mother tongue, but also two other European languages (Lasagabaster & López Beloqui, 2015). This goal, known as the 1+2 principle, resulted in the teaching of content classes in FLs across Europe (Eurydice, 2006). However, although CLIL instruction was initiated to promote multilingualism, CLIL programmes have predominantly taken place in English (San Isidro, 2018). This practice evidently entails that research has also focused on what Dalton-Puffer (2011) refers to as CEIL- Content and English Integrated Learning. For example, Cenoz et al. (2014: 257) note how “much, if not most, research on CLIL has been conducted by ESL/EFL scholars.” Similarly, Pérez et al. (2016: 485) have emphasized the lack of CLIL research on Languages Other Than English (LOTEs), even calling this dearth an “empirical vacuum” in the field. While numerous calls have been made to address this (e.g., Cenoz et al., 2014; Dalton-Puffer et al., 2010), there remains a clear lack of research investigating trilingual CLIL (Merino & Lasagabaster, 2018).

In a CLIL context, one area of language acquisition which has been found to be of great importance is vocabulary (Dalton-Puffer, 2008). This may be due to the fact that if vocabulary in content-based teaching is not understood, this will evidently affect the comprehension of the content which is being taught. In order to ascertain which vocabulary should be taught in this context, as well as to assess learners’ progress in this regard, Geoghegan and Agustín Llach (2023) have recently highlighted the usefulness of investigating learners’ lexical availability (LA). As outlined by Hernández-Muñoz et al. (2006: 730), LA is the measurement of “the ease with which a word can be generated as a member of a given category.” For example, in a category such as animals, it is likely that we have words such as “cat” or “dog” readily available to use.

One advantage of LA research is the use of the Lexical Availability Task (LAT), its primary data collection instrument, which has ensured methodological homogeneity (Geoghegan & Agustín Llach, 2023). Generally administered as a paper-and-pencil questionnaire, the LAT presents participants with a series of prompts, or semantic categories, and asks them to write down any words that come to their mind in response in a period of two minutes (Samper Hernández & Jiménez Catalán, 2014). Given the simplicity of the test, it can easily be replicated, using the same prompts that have been used in previous studies or focusing on specific lexical domains which are of key interest. In terms of CLIL teaching, LA could be used to target the specific content-related vocabulary of the subject at hand. However, research has yet to explore prompts which are actually related to the content to which students are exposed (Canga Alonso, 2017). In addition, while LA has been used extensively in the Spanish-speaking world, it has received relatively little attention in English (Ferreira et al., 2019), and research into other target languages (TLs) is virtually non-existent.

This present study aims to address these gaps by a) using content-related LA prompts to assess CLIL students and b) assessing CLIL students who are simultaneously taking CLIL

classes in two FLs. The following two sections provide an overview of the key research which has been carried out on these two areas, dealing first with vocabulary and LA in CLIL and then with LA research in different TLs.

1.1. Vocabulary and CLIL

It is well known that vocabulary plays an intrinsic role in learning an L2. Saville-Troike and Barto (2017), for example, even suggest that it may be the most vital level of L2 knowledge for learners. This is likely due to the fact that vocabulary is essential for communication, there is a clear awareness of its importance, and it also plays a critical role in the development of grammatical competence (Barcroft, 2004). Furthermore, the role of vocabulary is of central interest in a CLIL context, where technical, semi-technical, and general academic language, in particular, have been suggested to be positively affected (Dalton-Puffer, 2008). According to Baten et al. (2020), this is largely due to five main factors.

Firstly, a CLIL approach is thought to be conducive to learning vocabulary given the combination of both implicit and explicit learning conditions (Merikivi & Pietilä, 2014). For example, students take part in activities such as reading content-related texts, listening to instructions, and speaking to the teacher and other classmates (Tabuenca Cuevas & Alcaraz Mármol, 2014). In addition, CLIL students generally study vocabulary explicitly (Matiasek, 2005) in order to be able to understand the content of the class.

Secondly, CLIL teaching offers a more contextualised scenario wherein to practise the TL by fulfilling real purposes (Rodríguez Bonces, 2012). As Tabuenca Cuevas and Alcaraz Mármol (2014) point out, this more meaningful use of the language and greater involvement on the part of the students means that vocabulary will be more easily remembered.

Thirdly, one of the aims of CLIL is to focus on the acquisition of content, which entails less focus on the type of error correction which is more commonly seen in typical FL classrooms. This may help foster a more non-threatening atmosphere, where students are less afraid to use the language and less worried about making mistakes (Baten et al., 2020).

Fourthly, CLIL has been suggested to lend itself to a greater amount of student-teacher interaction, which can in turn lead to greater exposure to target vocabulary (Nikula, 2010). While this is a possibility, it should be noted that CLIL teaching does not necessarily ensure a greater amount of interaction (Baten et al., 2020). In fact, other research into this matter has actually found little interaction in CLIL classes. For example, Lo and Macaro (2015) found that, at least initially, teacher-student interaction in CLIL was actually rather monologic. More recently, comparing the same units in CLIL and non-CLIL settings, Alejo-González et al. (2021) noted no difference in the quantity of information received by students by means of interaction. Thus, while increased interaction is a potential benefit of CLIL instruction, whether or not there is a large quantity of interaction may vary greatly in different classrooms.

A final reason why CLIL may contribute to vocabulary development is that CLIL students are thought to have greater contact with the TL than non-CLIL students. This has been suggested in research by Sylvén (2019), where CLIL students also had a higher receptive vocabulary than their non-CLIL peers. Nonetheless, these advantages were found even prior to CLIL instruction. It thus appears in these studies that it is not CLIL teaching per se that leads to improvements in vocabulary, but that CLIL students simply tend to report greater extramural contact.

One area of vocabulary which is of particular interest in a CLIL context is LA. As pointed out by Geoghegan and Agustín Llach (2023), LA provides a useful means of assessing CLIL learners' vocabulary acquisition and also of determining the most suitable target vocabulary which needs to be taught in class. For example, students taking a subject such as economics could be tested using related prompts such as *Economy & Money*, in order to determine whether they have acquired the specific vocabulary which is needed to understand the content they are studying in their TL. Despite the usefulness of measuring CLIL students' LA, there remains a clear dearth of research investigating the LA of Spanish CLIL students. In addition, the two published papers on this topic deal with cohorts of students of very different ages (6th grade and 10th grade) and report very different results.

In a primary school context, Jiménez Catalán and Ojeda Alba (2009) compared the English LA of 42 CLIL and 44 non-CLIL 6th grade learners. Findings indicated that the non-CLIL learners had a significantly higher score on a language placement test than CLIL learners. Regarding LA, although there was no statistically significant difference, non-CLIL learners also produced a higher number of tokens than CLIL learners. In a secondary school context, Jiménez Catalán and Agustín Llach (2017) investigated the English LA of 8th grade (24 CLIL and 26 non-CLIL) and 10th grade (19 CLIL) learners. Notably, the 8th grade CLIL group had received the same number of hours of exposure to the TL as the 10th grade non-CLIL group, allowing the researchers to determine whether advantages could be attributed to time of exposure rather than the CLIL instruction itself. Findings revealed that the younger 8th grade CLIL group outperformed both their non-CLIL peers as well as the older, 10th grade non-CLIL group. This suggests that it is in fact the teaching context, rather than the consequent increase in time of exposure, which attributes to the advantages observed in LA.

While the above research indicates that CLIL teaching may be beneficial in terms of LA with older adolescent learners, more studies are evidently needed to determine whether this is in general the case in secondary CLIL contexts. In addition, research to date has yet to assess students in terms of the content vocabulary they are exposed to. In other words, research has included prompts such as *Animals*, which target vocabulary which may be completely unrelated to the content the learners are studying. As a result, researchers such as Canga Alonso (2017) have highlighted the need to include prompts which are actually related to the content of the students' CLIL classes, to better understand whether CLIL students really

acquire this vocabulary. The same researcher has also highlighted the need to include a language proficiency test alongside measuring LA, in order to establish the relationship between these two factors.

1.2. Token production in different TLs and across time

The analysis of token production in a LAT provides a useful method of assessing students' productive vocabulary in different languages. While there is a scarcity of research on this, some studies have indicated that English learners may produce more tokens than those studying other TLs. For example, Santos Díaz (2017) has found that while students produced the highest number of tokens in Spanish, their mother tongue, with a mean of 397.67, those studying English produced a mean of 261.67 and those studying French a mean of only 221.33. Šifrar Kalan (2014) similarly found slightly higher means for Slovenian university students learning English as compared to those learning Spanish, though there were evident similarities in the productivity of different prompts. Sandu and Konstantinidi (2021) also compared English and Spanish learners, namely upper-secondary school students studying Spanish in Romania and university students studying English in Spain. While both groups were reported to have similar language proficiencies in their chosen language, results again showed that the English L2 learners produced more words than the Spanish L2 learners. While these studies point to an English advantage, it should be noted that they analyse the responses of different cohorts of learners, and so any advantages observed could evidently be due to learner differences rather than the specific TL at hand.

In terms of differences in token production across time, research in a Spanish context concentrates solely on English, and has indicated that older, more advanced students tend to retrieve more tokens than younger students. For example, Agustín Llach and Fernández Fontecha (2014) investigated the responses of learners in both 6th grade and then three years later in 9th grade. Results showed that learners produced a statistically significant higher number of tokens in 9th grade than they had in 6th grade, indicating that their LA had improved over the three-year period as they aged and received more exposure to the TL. Similar results are reported by Jiménez Catalán and Fitzpatrick (2014), who compared 6th and 8th grade learners across nine different prompts. Findings revealed that the older group retrieved many more tokens (the number of words) and types (the number of unique words) than the younger group, and that the tokens produced also differed qualitatively. However, as both of the above studies compare young learners with adolescents, it could be the case that differences observed are due to age or cognitive level (Fernández Fontecha, 2015). Focusing solely on adolescent learners, Jiménez Catalán and Agustín Llach (2017), discussed above, compared the LA of teenagers in 8th and 10th grade in CLIL and non-CLIL education. Results revealed that the younger CLIL group outperformed both their non-CLIL peers, as well as the older non-CLIL group.

The above research suggests that there is an English advantage in LA and that while older, more proficient learners generally produce more tokens than younger learners, CLIL exposure may make a clear difference in the vocabulary that students produce. However, given that the research to date focuses on English, there is a need to compare different TLs in the same cohort of learners, so as to better understand the strengths and weaknesses of CLIL language-independently (Dalton-Puffer et al., 2010).

2. METHODOLOGY

2.1. Study design

As discussed above, calls have been made to assess both CLIL students' language level and their LA using content-related prompts, and to explore gains in CLIL in more than one TL. This study sought to address these gaps by investigating the quantitative differences in the language level and token production of adolescent CLIL students in English as compared to French in three different grades as well as across these grades. To this effect, it posed the following research questions:

1. RQ1 Are there quantitative differences in the language level of adolescent CLIL students in English as compared to French?
2. RQ2 Are there quantitative differences in the token production of adolescent CLIL students in English as compared to French?

To address these research questions, two data collections took place, separated by one full year. In the first data collection, 10th grade students only took part. The second data collection took place the following year with the same students, then in 11th grade, as well as 9th grade students (Table 1).

Table 1. Study Design.

Cross-sectional		
9 th grade	10 th grade	→ 11 th grade
Longitudinal		

While the original design of the study intended to assess 10th grade students on two occasions, in the same academic year, the ongoing pandemic made this impossible. Participating centres were thus contacted to rearrange the second data collection for a later date, when students had returned to school. A problematic consequence of this change was that when the second data collection took place, 10th grade students had transitioned to 11th grade and generally no longer took CLIL classes. In order to compensate for this issue, it was decided that 9th grade students

would also be assessed in the second data collection, so as to be able to compare groups which were taking CLIL classes in two different grades.

Using this design, data were analysed cross-sectionally, comparing 9th grade students with students with an additional year's CLIL experience in 10th grade, and longitudinally, comparing the 10th grade students who took part in both the first data collection and in the second data collection when they were in 11th grade. This resulted in a total of twelve data collections: six in English and six in French, each in 9th, 10th and 11th grade and each in the two participating schools. The tests were administered during the participants' normal FL class time: the English-focused tests were taken in their English-language class and the French-focused tests were taken in their French-language class. Consent forms were signed by the directors of the participating schools as well as each individual participant prior to each data collection.

2.2. Participants

Participants in the study included 91 Spanish native speakers from 9th, 10th and 11th grade. All participants attended one of two single-sex, semi-private sister schools which adopt a trilingual CLIL approach from early childhood education up to the end of compulsory secondary education, i.e., ages 3 to 16. The schools' plurilingual project states that each of the schools' three vehicular languages (Spanish, English, and French) should take up a third of the students' day. In this vein, in addition to their traditional language classes, students also took content classes in each of these languages, leading to an average of 10 hours of exposure to each language every week and an average of 350 hours each academic year. Regarding language level, the schools' objective is that by the time students finish secondary school, they will have achieved a B2 or C1 level in each TL. In 9th and 10th grade, language textbooks were aimed towards B2 level while in 11th grade they were aimed towards C1 level. Teachers are either native speakers or are accredited with a C1 level or higher and receive yearly linguistic and methodological training from the regional education authorities.

Of the 91 participants, 42 were in 9th grade and 49 were in 10th and/or 11th grade. There were 41 participants in 10th grade and 40 participants in 11th grade, with participants in 10th and 11th grade being largely the same ($n = 32$) (Table 2). As shown, there were seventeen students from 10th and 11th grade who took part in only one data collection (nine in 10th grade and eight in 11th grade). This was because while 10th grade is part of compulsory secondary education, 11th grade is the first year of high school. Thus, in some cases, students were no longer enrolled in the participating schools at the second data collection and others enrolled in the schools in 11th grade. There were also some male students who participated in the data collection for one language only. In 9th grade, while all 42 students participated in the French tests, a total of three students later decided not to take part in the English tests. Furthermore,

in 11th grade French became an elective subject for students attending the male school, and so four students no longer took French at this time.

Table 2. Participants.

Grade	Number of Participants
9 th Grade	42 (39 for English)
10 th Grade only	9
10 th + 11 th Grade	32 (30 for French)
11 th Grade only	8 (6 for French)
Total	91

In addition to their FL classes, students also studied several content classes in their TLs, including biology, physics and chemistry, economics, physical education and technology in English and geography and history in French. These classes varied depending on the grade and school at hand (Table 3).

Table 3. Participants' possible CLIL classes in each grade.

Grade	English	French
9 th Grade	Physics and Chemistry Biology Physical Education Technology	Geography and History
10 th Grade	Science / Economics Physical Education	Geography and History
11 th Grade	Economics (males only)	-

As shown, in these particular grades, CLIL classes were given predominantly in English, with up to four content classes as compared to just one in French. It should be borne in mind that while the schools' policies state that languages are evenly distributed across the day throughout primary and secondary education, towards the end of secondary education it appears that there was an increase in exposure to English.

While all 9th grade students took the same CLIL classes, differences arose from 10th grade, at which point students' core subject depended on whether they chose to study social sciences or science. For the former, students studied economics in English, while for the latter, male students studied physics and chemistry and female students studied biology, due to subject availability in the selected schools. In addition, while 11th grade students generally no longer took CLIL classes, male students in the social sciences track ($n = 10$) continued to study economics through English.

2.3. Instruments

2.3.1. *Language proficiency*

Two methods were used to address language level in this study. The first was a language level C-test, a text completion test whereby “proficiency can be measured via the rate of successful restorations of the missing message elements” (Grujić & Danilović 2012: 2). As noted by McManus (2011), C-tests have been repeatedly found to tap macro-level skills and processing and, consequently, have been long used as a method of testing language level (Daller et al., 2003). In this study, the English C-test was adopted from Daller and Phelan (2006) and included four texts from an online news website, each containing 20 gaps for a total of 80 gaps. The French C-test was adapted from McManus (2011) and again included four texts taken from a published newspaper article. While the original texts contained a total of 112 gaps with between 19 to 35 gaps per text, for the purposes of this study, the gaps in the final sentences of two texts were eliminated, resulting in an updated version with a total of 80 gaps with between 19 to 21 gaps per text. This was done to maintain a degree of consistency between the C-tests in each language in terms of length and time. In both C-tests, each text opened with one completed sentence. Subsequently, the second half of each second word was removed and replaced with a blank space. The C-tests were used to compare participants’ language level in English and French, e.g., comparing the English levels of 9th grade and 10th grade participants or comparing the French levels of participants in 10th and 11th grade. However, given that the English and French C-tests came from different sources, the participants’ scores in each C-test were evidently not comparable. In other words, a higher or lower score in one C-test would not necessarily indicate that a participant had a higher or lower level of English or French. As a result, a second means of addressing language level was also included, namely participants’ self-reported language levels. This was assessed in a motivation questionnaire which was distributed to the participants, as part of a section on personal information, such as the participant’s age, gender, nationality, and language learning background (for a detailed account of the motivation questionnaire and the results of this part of the study, see Geoghegan, 2024). Concerning their language learning background, participants were asked to state the FLs they knew, and whether they had a high, medium, or low level. These results were used to determine whether, according to the participants themselves, there were differences in TL proficiencies, e.g., comparing the English level and French level of participants in each grade.

2.3.2 *Lexical availability task*

Participants’ token production was assessed by means of two paper-and-pencil LATs. These two LATs contained five different prompts and were identical, except for the fact that one was administered in English and one in French. Each prompt was presented on a separate page and students were told to write down any words they thought of within in a two-minute period.

The task included two general prompts and three prompts which intended to tap into students' content-related vocabulary:

1. Animals / Les animaux
2. Food & Drink / La nourriture et les boissons
3. Sport & Physical Activities / Le sport et les activités physiques
4. Environment & Climate / L'environnement et le climat
5. Economy & Money / L'économie et l'argent

The first two general prompts, *Animals* and *Food & Drink*, were chosen as they had been previously used with Spanish adolescents and had been found to be particularly productive (see Canga Alonso, 2017; Fernández Orío & Jiménez Catalán, 2015). The three content-related prompts were included to specifically assess vocabulary used in the participants' content classes: *Sport & Physical Activities*, related to physical education which was studied through English; *Environment & Climate*, related to geography and history which was studied through French; and *Economy & Money*, related to economics which was studied through English. Although these prompts were chosen specifically to tap into the content-related vocabulary of the participants in this study, comparable prompts had been used in previous research, such as *Sports and Hobbies* (Agustín Llach & Fernández Fontecha, 2014) and *The Environment* and *The Economy* (Neilson Parada, 2016).

2.4. Analysis

Prior to the analysis, the data were prepared in order to run analyses using SPSS (Version 26). Responses for the C-test were scored according to the procedure used in Daller et al. (2003), accepting only exact solutions and awarding one point per correct answer, for a total score of 80. For the self-reported proficiencies, numerical values were assigned to the responses high/medium/low level in each language. Responses from the LAT were lemmatised following the procedure outlined by Jiménez Catalán and Agustín Llach (2017). This included correction of spelling errors; deletion of unintelligible words, repeated words, and proper nouns; lemmatisation of lexical phrases as one lexical unit; and changing plural words to the singular. For each prompt, tokens were counted using WordSmith Tools. Each participant received a numerical value for the total tokens produced in each prompt, and the average number of tokens produced overall was calculated.

To compare participants' self-reported proficiencies, Wilcoxon signed-rank tests were used to compare the level they had reported for English and French in each of the three grades. To determine whether there were differences in the participants' LA in English and French, paired samples t-tests were run to compare the languages in each grade under analysis. In all cases, analyses were run on each of the five individual prompts as well as the overall LAT. In

the case of the prompt *Animals* in French in 9th grade and in English in 10th grade, Wilcoxon signed-rank tests were used, given that the data were not normally distributed.

Regarding the differences across grades, participants were compared in terms of their performance on the language level C-tests and the LAT. To address differences in language level across grades, the results of the C-tests were analysed, due to the fact that each C-test could be compared from one grade to the next, rather than comparing the English and French tests with each other, as explained above. For both the cross-sectional and longitudinal data, as the data were normally distributed, paired samples t-tests were used to compare 9th and 10th grade students in each TL and to compare 10th and 11th grade students in each TL. To analyse LA across grades, independent samples t-tests were used to compare the cross-sectional differences while paired samples t-tests were used to compare the longitudinal differences. Again, for the prompt *Animals* in French in 9th grade and the prompt *Animals* in English in 10th grade, non-parametric tests were carried out, namely Mann-Whitney U tests for the cross-sectional analysis and Wilcoxon signed-rank tests for the longitudinal analysis.

3. RESULTS

3.1. Language proficiency

Research question one asked whether there were quantitative differences in the language level of adolescent CLIL students in English as compared to French. Regarding students' perceptions of their language level in English and French in each grade, results firstly indicated there were statistically significant differences between the two languages in 9th ($z = -2.80, p = .005$), 10th ($z = -2.98, p = .003$), and 11th grade ($z = -2.21, p = .027$). In each of the three grades, students reported a higher level of English (9th grade: $Mdn = 3.00, IQR = 2$; 10th grade: $Mdn = 2.00, IQR = 1$; 11th grade: $Mdn = 3.00, IQR = 2$) than French (9th grade: $Mdn = 3.00, IQR = 1$; 10th grade: $Mdn = 2.00, IQR = 0$; 11th grade $Mdn = 3.00, IQR = 4$).

Concerning performance on the language level C-tests across grades, results of the cross-sectional analysis found no statistically significant differences in either the English C-test or the French C-test between 9th grade and 10th grade, whereas the longitudinal analysis found statistically significant differences in both the English and French C-tests between 10th grade and 11th grade (Table 4).

These results indicate first and foremost that students across all three grades perceive their English language proficiency to be higher than their French language proficiency. In addition, results of the C-tests suggest that while there is little difference between the language level of 9th and 10th grade students in either language, the 10th grade students did improve their language level in both TLs from 10th to 11th grade.

Table 4. English and French C-tests across grades.

Cross Sectional (9 th and 10 th Grade)						
	<i>M</i>		<i>SD</i>		<i>t</i>	<i>p</i>
	9 th	10 th	9 th	10 th		
English	29.15	34.05	12.38	11.87	-1.83	.071
French	30.43	35.44	11.12	12.57	-1.92	.058
Longitudinal (10 th to 11 th Grade)						
	<i>M</i>		<i>SD</i>		<i>t</i>	<i>p</i>
	10 th	11 th	10 th	11 th		
English	33.78	36.91	12.84	12.61	-2.64	.013
French	35.97	42.60	12.57	9.04	-2.19	.036

3.2. Token production

Research question two asked whether there were quantitative differences in the token production of adolescent CLIL students in English as compared to French. Regarding students' token production in English and French in 9th, 10th and 11th grade, results revealed a clear advantage for English, as participants produced a statistically significant higher number of tokens in English than in French in all five prompts and overall, in all three grades (Table 5).

Table 5. Token Production in English and French.

9 th Grade						
<i>Prompt</i>	<i>M</i>		<i>SD</i>		<i>t/z</i>	<i>p</i>
	<i>English</i>	<i>French</i>	<i>English</i>	<i>French</i>		
<i>Animals*</i>	19.08	10.33	7.10	5.62	-4.91	<.001
<i>Food & Drink</i>	19.76	9.87	8.73	5.11	8.06	<.001
<i>Sport & Physical Activities</i>	14.68	9.82	5.96	4.02	5.52	<.001
<i>Environment & Climate</i>	17.31	10.21	7.63	5.28	6.72	<.001
<i>Economy & Money</i>	10.44	8.05	6.01	4.84	3.44	.001
<i>Mean Tokens</i>	16.24	9.67	6.24	4.02	8.69	<.001

Note. * = non-parametric test used, as the data were not normally distributed.

10 th Grade						
Prompt	M		SD		t/z	p
	English	French	English	French		
<i>Animals*</i>	18.22	8.54	4.83	3.47	-5.57	<.001
<i>Food & Drink</i>	21.61	10.80	5.29	5.28	12.93	<.001
<i>Sport & Physical Activities</i>	16.61	11.22	4.56	2.56	7.74	<.001
<i>Environment & Climate</i>	18.61	10.41	5.86	4.75	10.67	<.001
<i>Economy & Money</i>	13.95	10.41	4.79	3.74	4.51	<.001
<i>Mean Tokens</i>	17.80	10.27	4.00	3.05	15.16	<.001
11 th Grade						
Prompt	M		SD		t	p
	English	French	English	French		
<i>Animals</i>	20.19	11.06	5.85	3.84	11.27	<.001
<i>Food & Drink</i>	23.22	11.28	6.88	5.63	14.19	<.001
<i>Sport & Physical Activities</i>	17.42	10.17	4.77	3.12	9.46	<.001
<i>Environment & Climate</i>	18.69	12.08	7.47	5.99	7.60	<.001
<i>Economy & Money</i>	13.94	9.47	5.11	4.52	5.56	<.001
<i>Mean Tokens</i>	18.69	10.81	4.90	3.76	14.20	<.001

Note. * = non-parametric test used, as the data were not normally distributed.

Regarding students' token production across grades, results of the cross-sectional analysis found that while in English there was a statistically significant difference between students in 9th and 10th grade in just one prompt in English (*Economy & Money*) and in two prompts in French prompts (*Sport & Physical Activities* and *Economy & Money*) (Table 6).

Table 6. Cross-Sectional Token Production in English and French.

	Prompt	M		SD		t/z	p
		9 th	10 th	9 th	10 th		
	9 th → 10 th grade: English	<i>Animals*</i>	19.08	18.22	7.10	4.83	-.695
	<i>Food & Drink</i>	19.77	21.61	8.61	5.29	-1.14	.257
	<i>Sport & Physical Activities</i>	14.62	16.61	5.90	4.56	-1.69	.094
	<i>Environment & Climate</i>	17.31	18.61	7.63	5.86	-.85	.393
	<i>Economy & Money</i>	10.44	13.95	6.01	4.79	-2.89	.005
	<i>Mean Tokens English</i>	16.24	17.80	6.24	4.00	-1.32	.191

	<i>Prompt</i>	<i>M</i>		<i>SD</i>		<i>t / z</i>	<i>p</i>
		<i>9th</i>	<i>10th</i>	<i>9th</i>	<i>10th</i>		
9 th → 10 th grade: French	<i>Animals*</i>	10.33	8.54	5.62	3.47	-997	.319
	<i>Food & Drink</i>	9.83	10.80	4.93	5.28	-.86	.390
	<i>Sport & Physical Activities</i>	9.49	11.22	4.05	2.56	-2.31	0.24
	<i>Environment & Climate</i>	10.38	10.41	5.20	4.75	-.03	.976
	<i>Economy & Money</i>	7.86	10.41	5.00	3.74	-2.64	.010
	<i>Mean Tokens French</i>	9.59	10.27	3.92	3.05	-.88	.379

Note. * = non-parametric test used, as the data were not normally distributed.

Regarding the longitudinal analysis, results indicated that while in English there were no statistically significant differences in either overall token production or the five individual prompts, in French there were statistically significant differences in overall token production and in two of the five prompts (*Animals* and *Environment & Climate*) (Table 7).

Table 7. Longitudinal Token Production in English and French.

	<i>Prompt</i>	<i>M</i>		<i>SD</i>		<i>t / z</i>	<i>p</i>
		<i>10th</i>	<i>11th</i>	<i>10th</i>	<i>11th</i>		
10 th → 11 th grade: English	<i>Animals*</i>	18.63	20.25	4.84	5.81	-1.93	.053
	<i>Food & Drink</i>	22.16	23.06	5.45	7.38	-.954	.347
	<i>Sport & Physical Activities</i>	17.19	17.09	4.74	4.48	.152	.880
	<i>Environment & Climate</i>	18.59	19.31	6.34	7.69	-.833	.411
	<i>Economy & Money</i>	13.91	13.84	5.07	5.08	.098	.923
	<i>Mean Tokens English</i>	18.09	18.71	4.28	5.01	-1.37	.178
		<i>Prompt</i>	<i>M</i>		<i>SD</i>		<i>t</i>
<i>10th</i>			<i>11th</i>	<i>10th</i>	<i>11th</i>		
10 th → 11 th grade: French	<i>Animals</i>	8.77	11.40	3.42	3.86	-4.32	<.001
	<i>Food & Drink</i>	11.10	11.97	5.79	5.78	-1.31	.197
	<i>Sport & Physical Activities</i>	11.50	10.53	2.17	3.08	1.99	.056
	<i>Environment & Climate</i>	10.87	13.13	4.47	5.77	-3.36	.002
	<i>Economy & Money</i>	10.93	9.73	3.72	4.77	1.47	.152
	<i>Mean Tokens French</i>	10.63	11.35	3.06	3.78	-2.06	.048

Note. * = non-parametric test used, as the data were not normally distributed.

The analysis of token production thus indicates that, as in the case of language level, there appears to be an advantage for English, as students retrieve a statistically significant higher number of tokens than in French.

4. DISCUSSION

4.1. Language level and token production in English and French

The results above largely point to an English advantage, both in terms of language level and token production. These findings are unsurprising, given the focus on English in Spain, and the consequent higher amount of exposure that it receives. Although the participants in this study supposedly receive similar amounts of exposure to each TL (a third of each school day), in practice there is an evident emphasis on English in the grades under analysis. Furthermore, students' own perception of their language levels suggest that proficiencies are higher in English than in French, which is consistent with the results of the LAT and unsurprising if students receive more hours of exposure in English. Though previous research has scarcely addressed token production in different TLs, these findings are similar to the research discussed above by Šifrar Kalan (2014), Santos Díaz (2017), and Sandu and Konstantinidi (2021), which compared the LA of learners studying different languages and found that more responses are generally produced in English. Although these studies analyse the tokens of different cohorts of students, the present study found similar results when the same learners study two TLs simultaneously. The results also suggest that there is an evident relationship between language proficiency and LA, which is consistent with suggestions that more words are produced by more advanced learners (van Ginkel & van der Linden, 1996). As a result, it is very likely that the differences observed may not be attributable solely to the students' token production in each TL, but rather to their language proficiency in each one. It is thus imperative that, when comparing responses in English alongside other TLs, language level is taken into account. In other words, future research must endeavour to compare students with comparable language levels in order to determine whether advantages in LA are due to proficiency or to the specific TL.

4.2. Language level and token production in English and French across grades

Previous research on LA in English has indicated that older, more advanced students generally produce more tokens than younger students (e.g., Agustín Llach & Fernández Fontecha, 2014; Jiménez Catalán & Fitzpatrick, 2014). In the current study, however, there were clear differences in the two analyses across grades.

Firstly, the cross-sectional analysis revealed little difference between the 9th and 10th grade groups in terms of the language level C-tests and the overall LAT. However, in terms of the individual prompts, statistically significant differences were found in the prompts *Economy*

& *Money* in English and *Animals* and *Environment & Climate* in French. Given the fact that no difference was observed in language level or in other prompts, these advantages in 10th grade students may likely be due to the exposure received by the students rather than advantages in language proficiency. Regarding English, these results are extremely interesting as the prompt *Economy & Money* was related to economics, which students began to study in 10th grade. It thus appears that, although other prompts show no difference in the 9th and 10th grade groups, by studying economics in English, the older students improved their content-related vocabulary in the semantic field of economy. Regarding French, while the prompts *Sport & Physical Activities* and *Economy & Money* were intentionally included to investigate subject-specific vocabulary, they were related to subjects taken in English rather than French. However, it could be that participants have drawn from their English vocabulary to assist them in the completion of the French task. This would be consistent with previous research which has found that when students receive more exposure to a specific lexical domain, they tend to produce more tokens in related prompts (Canga Alonso, 2017; Fernández Orío & Jiménez Catalán, 2015).

Secondly, results of the longitudinal analysis indicated that differences depended to a large degree on the language at hand. While no differences were observed for the English LAT, there was an improvement in performance on the C-test. Meanwhile, for French there was an improvement in the prompts *Animals* and *Environment & Climate*, the overall LAT, and the C-test. Given that there was a statistically significant improvement in both C-tests, differences in the English and French LATs cannot be explained solely by differences in language level. A more likely explanation could be the fact that the number of tokens produced in English in both grades was generally much higher and so it could be that there is a ceiling effect at play, as there was less room for improvement. In French, on the other hand, as participants produced significantly fewer tokens in 10th grade, there was more room for improvement over the following year. This observation is consistent with previous research on productive vocabulary by Alejo González and Piquer Píriz (2016), who reported statistically significant growth in the productive vocabulary of lower-level secondary school students, but not for higher-level students. Other previous research by Agustín Llach and Fernández Fontecha (2014) and Jiménez Catalán and Fitzpatrick (2014) on age-related differences dealing specifically with LA has found that older learners tend to produce more words than younger learners. However, these studies compared 6th and 9th grade and 6th and 8th grade, respectively, and so the age gap at hand was quite a bit larger than that in the current study. Nonetheless, despite this shorter period of just one year, differences in token production were found for French, but not for English. This indicates that the specific TL, as well as the language proficiency which students have in the language under observation, may play a key role in longitudinal analysis of token production. This is a key area in need of further research, particularly given the lack of research that has been carried out on LOTEs.

5. CONCLUSION

This study addressed whether there were differences in the language level and token production of adolescent CLIL students in English and French, and whether these differences existed across grades. Results revealed that participants in all grades produced a statistically significant higher number of tokens in English than in French in all five prompts and overall, and also reported having a statistically significant higher language proficiency in English than in French. In terms of the differences across grades, results suggest that exposure to content-related vocabulary in CLIL may help to improve LA, and that students may draw from the exposure they receive in their English CLIL classes in the French task. In addition, the specific TL was found to play a key role across grades, with greater improvements in LA in French than English. While these findings offer some valuable insights, one potential issue in the study is the participants' language level in each TL, as participants both reported and demonstrated higher language levels in English than in French. To better compare LA in multilingual CLIL, it would be of the utmost value for future research to ensure that participants have similar proficiencies in each TL, incorporating a comparable language level assessment of the two TLs rather than resorting to students' self-reported levels. Considering Sylvén's (2019) research, which highlights that CLIL students tend to report greater extramural contact, it would also be extremely beneficial for future research to control for this variable. This would allow us to determine whether exposure to the FL outside of school has had an impact on the results. In addition, it would have been extremely interesting to conduct the longitudinal analysis over a longer period, to better understand the sequential development of the same learners in a CLIL context.

These findings indicate that language proficiency and exposure to content-related vocabulary play a key role in multilingual CLIL settings. It is extremely important that CLIL educators are made aware of this reality, so that they may endeavour to foster this multilingual learning environment.

REFERENCES

- Agustín Llach, M. P., & Fernández Fontecha, A. (2014). Lexical variation in learners' responses to cue words: The effect of gender. In R. M. Jiménez Catalán (Ed.), *Lexical availability in English and Spanish as a second language* (pp. 69-81). Springer.
- Alejo González, R., & Piquer Píriz, A. M. (2016). Measuring the productive vocabulary of secondary school CLIL students: Is Lex30 a valid test for low-level school learners? *Vigo International Journal of Applied Linguistics*, 13, 31-54.
- Alejo-González, R., Lucero, M., Schleppegrell, M., & Sánchez, A. (2021). Student-teacher interaction in CLIL and non-CLIL elementary education: A case study of verbal-pedagogical strategies and participation during read alouds in a Natural Science classroom. *Journal of Immersion and Content-Based Language Education*. <https://doi.org/10.1075/jicb.21005.ale>

- Barcroft, J. (2004). Second language vocabulary acquisition: A lexical input processing approach. *Foreign Language Annals*, 37(2), 200-208.
- Baten, K., van Hiel, S., & de Cuypere, L. (2020). Vocabulary development in a CLIL context: A comparison between French and English L2. *Studies in Second Language Learning and Teaching*, 10(2), 307-336.
- Canga Alonso, A. (2017). Spanish L1 speakers' and EFL learners' available lexicon. *Anuario de estudios filológicos*, 40, 5-23.
- Cenoz, J., Genesee, F., & Gorter, D. (2014). Critical analysis of CLIL: Taking stock and looking forward. *Applied Linguistics*, 35(3), 243-262.
- Daller, H., & Phelan, D. (2006). The C-test and TOEIC® as measures of students' progress in intensive short courses in EFL. In R. Grotjahn (Ed.), *Der C-Test: Theorie, Empirie, Anwendungen/The C-Test: Theory, empirical research, applications* (pp. 101-119). Peter Lang.
- Daller, H., van Hout, R., & Daller-Treffers, J. (2003). Lexical richness in the spontaneous speech of bilinguals. *Applied Linguistics*, 24, 197-222.
- Dalton-Puffer, C. (2008). Outcomes and processes in content and language integrated learning (CLIL): Current research from Europe. In W. Delanoy & L. Volkmann (Eds.), *Future perspectives for English language teaching* (pp. 139-157). Carl Winter.
- . (2011). Content-and-language integrated learning: From practice to principles? *Annual Review of applied linguistics*, 31, 182-204.
- Dalton-Puffer, C., Nikula, T., & Smit, U. (2010). *Language use and language learning in CLIL classrooms: Current findings and contentious issues*. John Benjamins.
- Eurydice (2006). *Content and Language Integrated Learning (CLIL) at School in Europe*. European Commission.
- Fernández Fontecha, A. (2015). Motivation and vocabulary breadth in CLIL and EFL contexts. Different age, same time of exposure. *Complutense Journal of English Studies*, 23, 79-96.
- Fernández Orío, S., & Jiménez Catalán, R. M. (2015). Lexical availability of EFL learners at the end of Spanish Secondary Education: The effect of language program and prompt. *ES Revista de Filología Inglesa*, 36, 103-127.
- Ferreira, R. A., Garrido Moscoso, J. I., & Guerra Rivera, A. (2019). Predictors of lexical availability in English as a second language. *Onomázein*, 46, 18-34.
- Geoghegan, L. (2024). Language Learning Motivation in Multilingual CLIL. *Porta Linguarum: Revista Internacional de Didáctica de las Lenguas Extranjeras*, 41, 137-151. <https://doi.org/10.30827/portalin.vi41.26620>
- Geoghegan, L. & Agustín-Llach, M. P. (2023). Theoretical Exploration of The Lexical Availability Task as A Tool for The Selection of Vocabulary to Be Taught in Class. *International Journal of TESOL Studies*, 5(2) 5-18. <https://doi.org/10.58304/ijts.20230202>
- Goris, J. A., Denessen, E. J. P. G., & Verhoeven, L. T. W. (2019). Effects of content and language integrated learning in Europe A systematic review of longitudinal experimental studies. *European Educational Research Journal*, 18(6), 675-698.
- Grujić, T., & Danilović, J. (2012). The Value of C-Tests in English Language Testing and Teaching. Paper presented at *Multidisciplinary Conferences Language, literature, values*. Faculty of Philosophy, University of Niš, Serbia, April.
- Hernández-Muñoz, N., Izura, C., & Ellis, A. W. (2006). Cognitive aspects of lexical availability. *European Journal of Cognitive Psychology*, 18(5), 730-755.
- Jiménez Catalán, R. M., & Agustín Llach, M. P. (2017). CLIL or time? Lexical profiles of CLIL and non-CLIL EFL learners. *System*, 66, 87-99.

- Jiménez Catalán, R. M., & Fitzpatrick, T. (2014). Frequency profiles of EFL learners' lexical availability. In R. M. Jiménez Catalán (Ed.), *Lexical Availability in English and Spanish as a Second Language* (pp. 83-100). Springer.
- Jiménez Catalán, R. M., & Ojeda Alba, J. (2009). Disponibilidad léxica en inglés como lengua extranjera en dos tipos de instrucción. *Lenguaje y Textos*, 30, 167-176.
- Lasagabaster, D., & López Beloqui, R. (2015). The impact of type of approach (CLIL versus EFL) and methodology (book-based versus project work) on motivation. *Porta Linguarum: Revista Internacional de Didáctica de las Lenguas Extranjeras*, 23, 41-57.
- Lo, Y., & Macaro, E. (2015). Getting used to content and language integrated learning: What can classroom interaction reveal? *The Language Learning Journal*, 43(3), 1-17.
- Matiasek, S. (2005). *English as the language of instruction in Austrian chemistry lessons: Instances of explicit language teaching*. Unpublished master's thesis, University of Vienna, Austria.
- McManus, K. (2011). *The development of aspect in a second language*. Unpublished doctoral dissertation, Newcastle University, England.
- Merikivi, R., & Pietilä, P. (2014). Vocabulary in CLIL and in mainstream education. *Journal of Language Teaching and Research*, 5(3), 487.
- Merino, J. A., & Lasagabaster, D. (2018). CLIL as a way to multilingualism. *International Journal of Bilingual Education and Bilingualism*, 21(1), 79-92.
- Neilson Parada, M. (2016). *Lexical Availability in Diaspora Spanish: A Crossgenerational Analysis of Chilean Swedes*. (Unpublished doctoral dissertation). University of Illinois at Chicago, Chicago, Illinois.
- Nikula, T. (2010). Effects of CLIL on a teacher's classroom language use. In C. Dalton-Puffer, T. Nikula, & U. Smit (Eds.), *Language use and language learning in CLIL classrooms* (pp. 105-124). John Benjamins.
- Pérez, A., Lorenzo, F., & Pavón, V. (2016). European bilingual models beyond linguafranca: key findings from CLIL French programs. *Language policy*, 15(4), 485-504.
- Rodríguez Bonces, J. (2012). Content and language integrated learning (CLIL): Considerations in the Colombian context. *GIST: Education and Learning Research Journal*, 6, 177-189.
- Ruiz de Zarobe, Y., & Lasagabaster, D. (2010). Introduction. In Y. Ruiz de Zarobe & D. Lasagabaster (Eds.), *CLIL in Spain: Implementation, results and teacher training* (pp. ix-xvii). Cambridge Scholars Publishing.
- Samper Hernández, M. & Jiménez Catalán, R. M. (2014). Researching lexical availability in L2: Some methodological issues. In R. M. Jiménez Catalán (Ed.), *Lexical availability in English and Spanish as a second language* (pp. 189-205). Springer.
- San Isidro, X. (2018). Innovations and challenges in CLIL implementation in Europe. *Theory into Practice*, 57(3), 185-195.
- Sandu, B. M., & Konstantinidi, K. (2021). Impact of overseas experience on lexical availability in english and spanish as foreign languages. In A. B. Barragán Martín, M. M. Molero Jurado, Á. Martos Martíne, M. M. Simón Márquez, J. J. Gázquez Linares, M. C. Pérez Fuentes (Eds.), *Innovación docente e investigación en arte y humanidades: nuevos enfoques en la metodología docente* (pp. 1049-1058). Dykinson.
- Santos Díaz, I. C. (2017). Selección del léxico disponible: propuesta metodológica con fines didácticos. *Porta Linguarum: Revista Internacional de Didáctica de las Lenguas Extranjeras*, 27, 122-139.
- Saville-Troike, M., & Barto, K. (2017). *Introducing Second Language Acquisition*. (3rd ed.). Cambridge University Press.

- Šifrar Kalan, M. (2014). Slovene students' lexical availability in English and Spanish. In R. M. Jiménez Catalán (Ed.), *Lexical Availability in English and Spanish as a Second Language* (pp. 125-138). Springer.
- Sylvén, L. K. (2019). Extramural English. In L. K. Sylvén (Ed.), *Investigating Content and Language Integrated Learning: Insights from Swedish High Schools* (pp.152-167). Multilingual Matters.
- Tabuenca Cuevas, M. F., & Alcaraz Mármol, G. (2014). From EFL to CLIL teaching and learning vocabulary in the primary classroom. In J. R. Calvo-Ferrer & M. A. Campos-Pardillos (Eds.), *Investigating lexis vocabulary teaching, ESP, lexicography and lexical innovation* (pp. 27-37). Cambridge Scholars Publishing.
- van Ginkel, C. I., & van der Linden, E. H. (1996). Word associations in foreign language learning and foreign language loss. In K. Sajavaara & C. Fairweather (Eds.), *Approaches to second language acquisition* (pp. 25-35). University of Jyväskylä.