DOES CLIL EXPOSURE AFFECT THE ACQUISITION OF REFERENCE IN NARRATIVES? A CORPUS-BASED STUDY OF L2-ENGLISH

¿AFECTA EL AICLE A LA ADQUISICIÓN DE EXPRESIONES REFERENCIALES? UN ESTUDIO DE CORPUS DE INGLÉS COMO L2

Teresa Quesada
Universidad de Granada, Spain
teresaquesada@ugr.es

A great amount of literature in the Spanish educational context investigates the possible benefits of Content and Language Integrated Learning (CLIL) instruction. Some studies focus on the benefits of CLIL in general proficiency, while others focus on specific areas of language (Ruiz de Zarobe, 2011). These studies show that CLIL seems to be beneficial for general proficiency (though there is counterevidence), but the benefits are not so evident for specific linguistic areas. The use of referring expressions (REs) is a specific linguistic area well investigated in Second Language Acquisition Research but still not explored in CLIL settings. This study, thus, investigates this phenomenon from the point of view of CLIL instruction and tries to add some insights to the field of SLA. Therefore, the main goal of this study is to investigate the use of REs in discourse and the factors that constraint their use and determine whether their acquisition depends on the exposure to CLIL. In order to do

so, we use a corpus-based methodology and analyse written narratives comparing CLIL and non-CLIL learners. Results show that i) the use of REs is constrained by several factors and ii) CLIL instruction may be beneficial for the use of REs at early stages but CLIL and non-CLIL learners end up behaving similarly at intermediate stages and their production is still far from the native norm; that is, longer exposure to the L2 through CLIL instruction does not ultimately pose benefits at the syntax-discourse interface.

Key words: *CLIL, learner corpora, narratives, referring expressions, Second Language Acquisition (SLA)*

Gran parte de la literatura en el contexto educativo español investiga los posibles beneficios del Aprendizaje Integrado de Contenidos y Lenguas Extranjeras (AICLE). Algunos estudios se centran en los beneficios de AICLE en cuanto al nivel de competencia en la L2 en general, mientras que otros se centran en áreas del lenguaje específicas. Estos estudios muestran que el AICLE es beneficioso en cuanto al nivel de competencia alcanzado, pero no lo es tanto para áreas específicas del lenguaje. El uso de expresiones referenciales es un área del lenguaje concreta muy estudiada en la Adquisición de Segundas Lenguas, pero no en AICLE. Por tanto, este estudio investiga este fenómeno desde el punto de vista del AICLE. El principal objetivo del estudio es investigar el uso de expresiones referenciales en el discurso y los factores que determinan su uso y explorar si su adquisición depende de la exposición al AICLE. Para ello, utilizamos una metodología de corpus y analizamos narraciones escritas comparando alumnos que reciben AICLE y los que no. Los resultados demuestran que i) el uso de expresiones referenciales está determinado por diversos factores y ii) AICLE puede ser beneficioso para el uso de expresiones referenciales en niveles iniciales pero los alumnos de AICLE y no AICLE en niveles intermedios se comportan igual y su producción todavía difiere de la de los nativos; esto es, una mayor exposición a la L2 a través del AICLE no supone unos beneficios para interfaz sintaxis-discurso.

Palabras clave: *AICLE, corpus de aprendices, narrativas, expresiones referenciales, Adquisición de Segundas Lenguas (ASL)*

1. Introduction

The European Union initiative of raising multilingual societies led to the implementation of a wide range of bilingual education programmes. One of those programmes, which was created in Europe, became known as Content and Language Integrated Learning (CLIL) (Marsh, 2002). Spain followed the European initiative and bilingual programmes, particularly CLIL programmes, were implemented in several regions, although these programmes adopted different shapes (Martínez Adrián, 2011; Ruiz de Zarobe & Jiménez Catalán, 2009). Since the implementation of CLIL programmes in Spain, there has been a considerable amount of research trying to ascertain the general or specific benefits that these programmes can bring to the process of learning an L2 or L3. Several studies have shown the purported benefits of CLIL programmes in general proficiency and vocabulary (Martínez Adrián, 2011; Ruiz de Zarobe, 2011 for an overview), although other studies disagree with these supposed benefits (Bruton, 2011a, 2011b; Cenoz, Genesee, & Gorter, 2014 for an overview). When it comes to specific areas of language (i.e. phonetics, morphology, syntax, etc.), it has been shown that the benefits of CLIL programmes are not so evident (García Mayo & Villarreal Olaizola, 2011; Martínez-Adrián & Gutiérrez-Mangado, 2015; Martínez-Adrián & Gutiérrez-Mangado, 2009), but it is also claimed that further research is needed in this area (Cenoz et al., 2014; Martínez Adrián, 2011; Ruiz de Zarobe, 2011). Additionally, it is worth mentioning that there is a lack of research among CLIL studies at the syntax-discourse area, a widely studied phenomenon in Second Language Acquisition (SLA).

The aim of this paper, thus, is to carry out a pilot study and explore the syntax-discourse interface and the possible (beneficial) influence that longer exposure (through CLIL instruction) may have on this area. In order to do so, we focus on written narratives produced by L1 Spanish – L2 English learners attending CLIL and non-CLIL programmes and investigate i) general features of written production (in line with previous CLIL studies on production) and ii) specific features that constrain the use of referring expressions (REs) in the discourse (i.e. a phenomenon at the syntax-discourse interface). Additionally, we carried out this study following a corpus-based methodology and a fine-grained and linguistically-informed annotation system.

2. Background

The production of cohesive discourse is achieved through the use of a particular referring expression (RE) (i.e. noun phrase (NP), overt or null forms) at a given point in the discourse (Crosthwaite, 2013). The use of referring expressions (REs) in discourse to achieve coherence has been widely studied and it has been shown that the choice of a RE is constrained by several factors like: type of language, distance, information status, or saliency (for a more detailed explanation see: Crosthwaite, 2013; Gelormini-Lezama & Almor, 2011; Hendriks, Koster, & Hoeks, 2014; Lozano, 2016). Among those factors, the type of language and the information status will be of interest for the present study.

When studying the use of REs in discourse, it is very important to consider the type of language as it partly determines the choice of the RE. There is a major distinction between pro-drop and non-pro-drop languages (White, 2009). Pro-drop languages such as Spanish, Italian, or Greek allow the use of null pronouns as in (1), while non-pro-drop languages like English, German or French, typically require an overt subject and the use of null forms is restricted to specific contexts as in (2). As for the information status, Lozano (2009) states the distinction between topic continuity and topic shift contexts. In topic continuity contexts, the RE refers to the previous antecedent in the discourse and less information is required, while in topic shift contexts, there is a change in the character that is being mentioned and more information is required. Example (3) shows the alternation between topic continuity and topic shift contexts and consequently the alternation between NP, overt and null forms depending on the context. As mentioned above, the literature shows that there are many different factors affecting the choice of the RE; however, providing a complete account of all those factors is beyond the scope of this study and, therefore, we mainly focus on the type of language and the information status.

(1)   *El niño se fue a dormir. Al día siguiente Ø vio que la rana no estaba en la habitación. Ø decidió ir al bosque con su perro y buscar a la rana allí.* (The boy went to sleep. Next day, Ø* noticed that the frog wasn't in the room. Ø*decided to go to the forest with his dog and Ø looked for the frog there)

(2)   The boy went to sleep. Next day, he noticed that the frog wasn't in the room. He decided to go to the forest with his dog and Ø looked for the frog there.

(3)   One day, a small boy decided to catch a frog and Ø keep it in a jar. The frog was not a fan of this living arrangement and Ø escaped while the boy and his dog slept. In the morning the child and the dog went to look in the forest for the frog. They startle a deer and Ø are pushed into the river… [EN_WR_20_USA_TQC_AMS.txt]

The study of REs in narratives has received great attention in the Second Language Acquisition (SLA) literature. The literature shows that the use of REs in discourse seems to be problematic for L2 learners (L2ers). Initially, special attention was paid to the acquisition of RE in pro-drop languages (White, 2009), as it seemed to be particularly problematic, but the literature also shows evidence that L2ers' use of RE in non-pro-drop languages like English is also challenging (H. Hendriks, 2003; Kang, 2004; Pladevall Ballester, 2013; Prentza, 2014; Ryan, 2015).

Most L2 English studies that focus on the use of REs in discourse have studied oral production using a wide range tasks and participants with different L1s (Crosthwaite, 2011; H. Hendriks, 2003; Kang, 2004; Leclercq & Lenart, 2013; Ryan, 2015). Overall, it is generally agreed that L2 English learners (most of them at intermediate levels) show difficulties in the correct use of REs and tend to differ from native speaker production by being overexplicit (i.e. using fuller forms and being more specific than required). Therefore, longer time of exposure would be needed for these learners in order to achieve a native-like use of REs in discourse. This fact led us to the main idea of the present study. This study investigates the use of REs in discourse in learners attending to CLIL and non-CLIL programmes. If the use of REs depends on the amount of exposure, we question whether CLIL exposure will influence the acquisition of REs. If so, CLIL learners are expected to perform better than their non-CLIL counterparts.

3. CLIL Research and CLIL on Specific Areas of Language

It is well-known that CLIL is an 'umbrella' term adopted for some educational programmes that integrate language and content for learning

the L2, although it is also claimed that these programmes adopt different shapes and it is difficult to make generalizations about CLIL programmes (Cenoz et al., 2014). Therefore, it is important to define the boundaries of CLIL when we examine CLIL research and to what extent results can be extrapolated. If we focus on the Spanish context, as the present study does, CLIL was implemented in several regions but the linguistic context is different in each region and its implementation will differ. Spain offers a variety of language contexts (Martínez Adrián, 2011; Ruiz de Zarobe & Jiménez Catalán, 2009) as there are monolingual and bilingual communities: the former offers English as the L2, while the latter offers English as the L3. Therefore, even within the same country, there are different scenarios, which should be acknowledged when considering research outcomes.

Having this in mind, CLIL research in Spain has so far shown varied results depending on the subject of study. If we focus on general proficiency, it is generally claimed that CLIL programmes are beneficial as CLIL learners tend to outperform non-CLIL learners (Lasagabaster, 2008; Martínez Adrián, 2011; Pérez Cañado, 2018; Ruiz de Zarobe, 2011), although there is also counterevidence for these results (see Bruton, 2011a, 2011b for an overview). Other area that seems to be enhanced by CLIL instruction is vocabulary knowledge, in particular receptive vocabulary (Jiménez Catalán, Ruiz de Zarobe, & Cenoz, 2006; Ruiz de Zarobe & Jiménez Catalán, 2009). When it comes to specific linguistic areas, more research is still needed but it has been mostly shown that CLIL programmes do not affect the development of syntax or morphology and focus on form would be required (García Mayo & Villarreal Olaizola, 2011; Martínez Adrián & Gutiérrez Mangado, 2015; Martínez-Adrián & Gutiérrez-Mangado, 2009; Villarreal Olaizola & García Mayo, 2009).

Focusing on the study of REs in discourse, there is a lack of research in the CLIL literature in this area. Several studies consider narratives in oral production (Adrián & Mangado, 2015; Lázaro, 2012; Lázaro & García, 2012; Llinares & Whittaker, 2007; Martínez-Adrián & Gutiérrez-Mangado, 2009) but these studies focus on specific syntactic aspects or on the narrative as a whole in terms of accuracy, syntactic complexity, lexical richness, etc., and not so much in the use of REs in discourse and the factors that constrain their use. Some of these studies

address the use of pronouns in general terms including all types of pronouns. Also, they only focus on the presence or absence of pronouns regardless of the factors that constrain their use. Lázaro (2012) and Lázaro and García-Mayo (2012) report that CLIL learners use a wider variety of pronominal forms and their use is more correct than the use and variety of non-CLIL learners, although the type of pronominal forms used and the contexts in which they occur are not specified. Martínez-Adrián and Gutierrez-Mangado (2009) focus on the use of incorrect null subjects and show that the production of incorrect null subjects is lower in the CLIL group but there are no statistically significant differences compared to the non-CLIL group. Recently, Gutiérrez-Mangado & Martínez-Adrián (2018) have focused on the interfaces but they do not investigate the syntax-discourse interface, the phenomenon under investigation in the present study. By contrast, they focus on the syntax-morphology interface and the syntax-semantic-discourse interface and show that there are no differences between CLIL and non-CLIL learners at the syntax-morphology interface, although CLIL instruction poses benefits at the syntax-semantics-morphology interface. To the best of our knowledge, there are no studies addressing the use of REs in discourse (a phenomenon at the syntax-discourse interface) following a fine-grained corpus-based method in the CLIL literature.

The purpose of this study, thus, is to investigate the use of REs in discourse comparing learners who attend CLIL and non-CLIL programmes. In the section above we explained that the use of REs is particularly challenging for L2ers. If the correct use of REs in discourse depends on the amount of input received, we believe that greater amount of input will lead to a more native-like use of REs. In this study, we compare CLIL and non-CLIL learners in order to see if a greater input to the language (in these type of contexts) affects the use of REs and lead to a better performance.

4. Research Questions and Predictions

RQ1: Are there differences in terms of length and complexity in the narratives of learners attending CLIL and non-CLIL programmes? Do they show differences compared to native speaker production?

H1: As previous literature shows, CLIL learners tend to outperform non-CLIL learners when it comes to general competence. The prediction for this general question is that CLIL learners will show longer (in terms of word length) and more complex (in term of sentence variety) narratives than their non-CLIL counterparts. If we compare both groups of learners (CLIL and non-CLIL) to native speakers, we expect that learners will differ from native speakers as they are still in the process of acquiring the L2 and their narratives will be probably simpler.

RQ2: What referring expressions (REs) do CLIL and non-CLIL learners use in their narratives compared to the use of native speakers? Are there differences between groups of learners according to proficiency level?

H2: As English is a non-null subject language, fuller forms (overt and NP) rather than null forms are expected in both native speakers and learners; however, we expect that learners will not use REs in a native-like way since the use of REs in discourse is challenging for L2ers (see 2. Background). In particular, we expect that both groups of learners (CLIL and non-CLIL) will differ from native speakers but the CLIL group is expected to be closer to the native norm as a result of a longer exposure to the L2.

RQ3: Does the character type (i.e. the character who is mentioned at a given point) constrain the use of REs? Are there differences between learners and native speakers and between CLIL and non-CLIL learners?

H3: Previous literature addressing this issue and following a similar methodology (Kang, 2004) shows that the choice of the RE varies depending on the character type and additionally this choice also varies among learners and native speakers. Therefore, we expect that the choice of the RE will vary depending on the character. Also, we expect to find differences between native speakers and learners in general and between CLIL learners and non-CLIL learners in particular so that the former will show a more native-like performance.

RQ4: Does information status influence the choice of the RE? Are there differences between learners and native speakers and between CLIL

and non-CLIL learners? Do CLIL learners produce a more cohesive discourse?

H4: As shown in examples (3), the information status context determines the choice of the RE. Here, we will focus on the cohesion of narrative and therefore in the continuation of topic. Thus, we expect that in topic continuity contexts less full forms will be produced (i.e. overt and null forms). We also expect a difference between native speakers and learners as the literature shows (see section 2. Background) that the use of RE may be difficult to be acquired and these learners are expected to be in the process of acquiring it. However, as CLIL learners are more exposed to the L2, we expect a better performance and a more cohesive discourse.

5. Method

5.1. Corpus and Participants

This study follows a corpus-based methodology. The corpus employed for the study was the Corpus of English as a Foreign Language (COREFL) (see Lozano, Díaz-Negrillo, & Callies, forthcoming for a complete overview). To summarize, COREFL started to be compiled in 2012 at the University of Granada and is still under compilation. The core component of COREFL is that it offers data from L2 English learners and native English speakers. However, COREFL has been expanded and different variables were introduced (i.e. L1, age, task type, mode, etc.) and therefore COREFL includes several subcorpora. One of those subcorpora contains written data from secondary school students who attended CLIL and non-CLIL programmes. A sample of this subcorpus, together with a sample of the English native speakers subcorpus (for the purpose of comparison), were selected for the present study. The tasks that these two subcorpora comprise are explained below (section 5.2).

The data collection procedure for the COREFL subcorpus of secondary students used for the present study was the following. The data was collected by Master students at the University of Granada as part of their MA Thesis. These students went to several secondary schools (following CLIL and non-CLIL programmes) in Granada and Jaén. As there is a wide range of schools, the subjects that the CLIL

groups received as part of the CLIL programme are also varied: social sciences, natural sciences, philosophy, art, integrated project, ethic and maths.

Participants from this subcorpus were chosen for the present study according to the following criteria. We first selected CLIL learners who had been attending a CLIL programme as long as possible. Then, we grouped those learners according to proficiency level (from A1 to B2) and tried to have a similar number of participants per proficiency level; also, we tried (as far as possible) to make sure that learners attended the same

Group name	N	Proficiency level	Mean chronological age	Actual course
CLIL	11	A1	14.5	2º ESO (n=5) 3º ESO (n=4) 4º ESO (n=2)
	11	A2	16	4ºESO (n=3) 1ºBACH (n=8)
	11	B1	16.2	1ºBACH (n=11)
	11	B2	16.3	4ºESO (n=1) 1ºBACH (n=10)
Non-CLIL	11	A1	14.2	2º ESO (n=4) 3º ESO (n=7)
	11	A2	16.1	4ºESO (n=4) 1ºBACH (n=7)
	11	B1	16.1	4ºESO (n=7) 1ºBACH (n=4)
	9	B2	15.4	4ºESO (n=5) 1ºBACH (n=4)
Native speakers	11	NA	20.4	University

Table 1. Participants

grade, although this was not always possible. Finally, we selected the non-CLIL learners who matched with the CLIL participants in terms of proficiency level and grade. All in all, the participants for this study include 44 learners attending to a CLIL programme (at 4 proficiency levels), 42 learners attending a non-CLIL programme (at 4 proficiency levels) and 11 native English speakers at university level as Table 1 below shows. Our corpus included many different variables and CLIL and non-CLIL learners could be grouped in different ways. For the purpose of this study, we considered appropriate to compare learners who are at the same proficiency level (regardless of the course level) in order to investigate the phenomena. However, as we have plenty of metadata about our participants, we can group them differently in the future for further comparisons.

5.2. Materials and Procedure

The task employed to elicit the data was *Frog, where are you?* (Mayer, 1969). Students at secondary school did the task as a classroom activity. They were presented with a selection of pictures from this story and they had to retell the story as a writing task. Then, the texts were transcribed to txt format. Additionally, the students completed a learning background questionnaire in order to have learners' biodata and an English Placement Test (Cambridge University Press, 2010) to control for the proficiency level. This information was incorporated to each txt file. Native English speakers did the same task (apart from the proficiency level test) but the data was collected online via google forms and then converted into txt files.

Once the sample from the corpus was gathered, the tagging procedure was carried out with UAM Software Tool (O'Donnell, 2008). This software is convenient because it does not only allow to create a tagset with the desired features for your project, but it also offers a fine analysis with different triangulations and some statistical tests. Therefore, first a tagset (see section 5.3) was created. Then, the texts were manually tagged following the tagset shown in Figure 1 below. For this project, we tagged all third person REs in subject position as the literature shows that they are particularly problematic (see Lozano 2016). Finally, the analysis was carried out in the software.
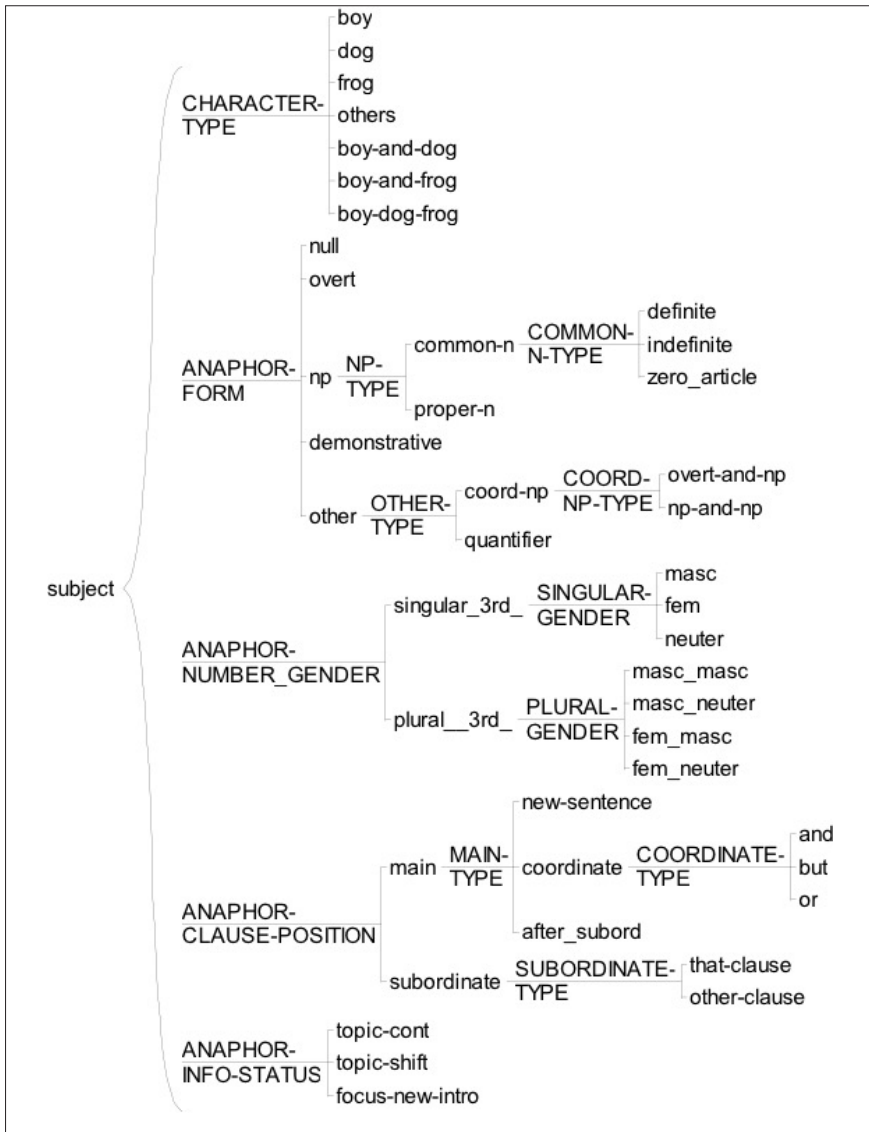
Figure 1. Tagset

5.3. Tagset

Figure 1 shows the tagset employed for the present study. Each 3th person RE in subject position was assigned the following tags. The system *character type* indicates the character to which the RE refers (boy, dog, frog, other characters, or different combinations of the main characters). The system *anaphor form* indicates the form of the RE (null, overt, Noun Phrase (NP), demonstrative, or other). The *number and gender* of the RE were indicated in the following system (we considered singular and plural RE and masculine, feminine and neuter genders). The *type of sentence* system indicates the type of sentence (i.e. main or subordinate) in which the RE appears. Additionally, the *main sentence* tag was divided into *new sentence*, *coordinate* and *after subordinate*: *new sentence* and *after subordinate* were considered instances of *main sentences* and they were distinguished from *coordinate* sentences for the analysis. We also classified the subordinate type into *that-clause* and *other-clause* but they were not further analysed in the present study. Finally, the *information status* system indicates the context in which the RE is produced (i.e. topic continuity, topic shift, or focus new introduction).

5.4. Data Analysis

The analysis was carried out with the same software employed for the tagging procedure (UAM Corpus Tool). This software allowed to make several comparisons between and within groups, exploring all the features incorporated in the tagset. The software provides the raw frequencies and the corresponding percentages; additionally, it provides the chi-square (χ^2) for each comparison and its approximate statistical significant levels (p-value): weak significance (90%: p=0.1); medium significance (95%; p=0.05) and high significant (99%; p=0.01). We consider significant differences for the present study when we got medium or high significance (95% and 99%) so the results section below only includes such significance levels. Additionally, one-way Analysis of Variance (ANOVA) and Tukey post hoc tests were carried out in the software R to show results in 6.1. Additionally, it is important to mention that even though 3th person singular and plural REs were tagged, only 3th person singular REs were considered for the analysis in this study.
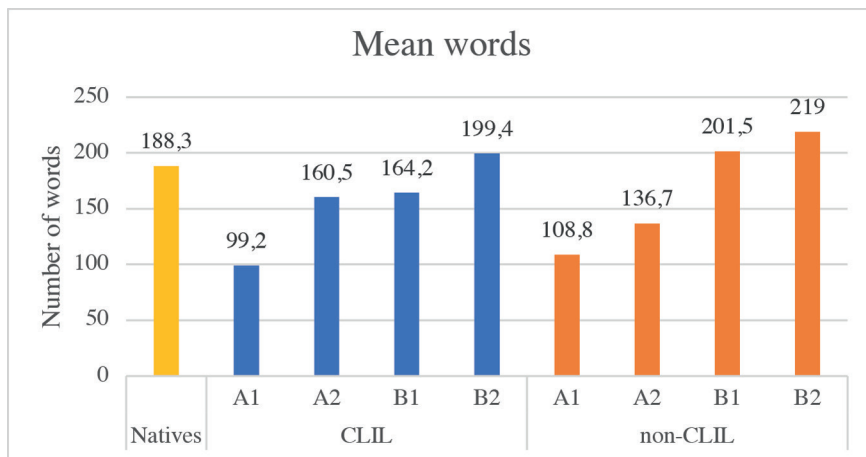
Figure 2. Mean words per group and across proficiency level

6. Results

6.1. RQ1. Word Length and Complexity of Narrative

The first RQ addresses narratives' length and syntactic complexity. Figure 2 shows the mean production of word tokens in the narratives of learners and native speakers. There is a statistically significant difference in the mean of words between groups as determined by one-way ANOVA (F $(2,27)$ = 6.805, p = .000000634). A Tukey post hoc test revealed there were statistically significant differences between: native and CLIL-A1 groups (p=.0044621); native and non-CLIL-A1 groups (p=.0177599); CLIL A1 and CLIL-B2 groups (p=.0007746); non-CLIL-A1 and non-CLIL-B1 groups (p=.0025573); non-CLIL-A1 and non-CLIL-B2 groups (p=.0003747); and non-CLIL-A2 and non-CLIL-B2 groups (p=.0214498). H1 predicted differences between learners and native speakers and between CLIL and non-CLIL groups. Differences between native speakers and both A1 groups (CLIL and non-CLIL) are confirmed but there are no significant differences between native speakers and the other groups. When it comes to differences between learner groups, some differences were found between CLIL groups at different proficiency levels and between non-CLIL groups at different proficiency levels. Importantly, there were no significant differences between CLIL and

non-CLIL groups at any proficiency level, which goes against our prediction and therefore it is shown that longer exposure does not affect the length of their narratives.

When it comes to the type of sentences they produce, Table 2 shows the production of main and subordinate sentences. There is a higher production of main sentences in all groups; however, there are significant differences between native speakers and CLIL-A1 groups , native speakers and non-CLIL-A1, native speakers and non-CLIL-A2, and CLIL-A2 and non-CLIL-A2 groups. This shows that both CLIL and non-CLIL beginner learners show a lower production of subordinate sentences compared to native speakers and the CLIL-A2 group significantly produces more subordinate sentences than the non-CLIL-A2 group.

Sentence Type		Main	Subordinate
CLIL	A1	86.9 % (126/145)	13.1 % (19/145)
	A2	77.8 % (168/216)	22.2 % (48/216)
	B1	79.9 % (191/239)	20.1 % (48/239)
	B2	76.6 % (193/252)	23.4 % (59/252)
non-CLIL	A1	93.4 % (142/152)	6.6 % (10/152)
	A2	85.9 % (171/199)	14.1 % (28/199)
	B1	82.4 % (215/261)	17.6 % (46/261)
	B2	76.7 % (191/249)	23.3 % (58/249)
Native speakers		77.9 % (183/235)	22.1 % (52/235)

Table 2. Production of sentence type by group across proficiency level

The tag *main type sentence* offered two different tags (see 5.3. Tagset above): *main sentence* (*new sentence* and *after subordinate*) and *coordinate* sentence. We further analysed if there were differences in the production of these two types of sentences. Overall, Table 3 shows that there is a higher production of main type sentences across groups. There is just a significant difference between CLIL-A1 and non-CLIL-A1 groups, which reveals that the CLIL-A1 group produces more coordinate

sentences than the non-CLIL-A1 group and therefore a more cohesive discourse. However, both CLIL and non-CLIL groups at subsequent proficiency levels behave similarly as there are no differences between them.

Sentence Type		Main (new sentence)	Coordinate
CLIL	A1	60.3% (76/126)	39.7 % (50/126)
	A2	60.1 % (101/168)	39.9 % (67/168)
	B1	67.5 % (129/191)	32.5 % (62/191)
	B2	75.7 % (146/193)	24.3 % (47/193)
non-CLIL	A1	77.5 % (110/142)	22.5 % (32/142)
	A2	68.4 % (117/171)	31.6 % (54/171)
	B1	63.3 % (136/215)	36.7 % (79/215)
	B2	70.7 % (135/191)	29.3 % (56/191)
Native speakers		70 % (128/183)	30 % (55/183)

Table 3. Production of main sentence type by group across proficiency level

6.2. RQ2. Overall Use of REs

H2 predicted that there would be a higher production of overt and NP forms in all the narratives (both learners and native speakers), but also some differences between groups were predicted. In this section, we will firstly focus on the native production and then we will see CLIL and non-CLIL groups production across proficiency levels plus the differences between them. As Table 4 and 5 show, native English speakers use more NP forms (49.5%) in their narratives, followed by overt (35.1%) and null (15.4%) forms. When it comes to learners, Table 4 shows that CLIL-A1 learners behave as native speakers using more NP (55.2%) forms, followed by overt (35.3%) and null (9.5%) forms. By contrast, CLIL-A2 and CLIL-B1 groups behave differently opting for a higher production of overt (A2: 49.4%; B1: 44.3%) forms, followed by NP (A2: 39.7%; B1:

50%) and null (A2: 10.9%; B1: 5.7%) forms. Then, the CLIL-B2 group again performs in a native-like manner showing a higher production of NP (50.8%) forms, followed by overt (44%) and null (5.2%) forms. As for the non-CLIL groups, Table 5 shows that the non-CLIL-A1 and the non-CLIL-A2 groups start showing a higher production of NP (non-CLIL-A1: 81.5%; non-CLIL-A2: 59.6%) forms, but to a greater extent than native speakers (especially the non-CLIL-A1 group), followed by overt (non-CLIL-A1: 12.9%; non-CLIL-A2: 30.4%) and null (non-CLIL-A1: 5.6%; non-CLIL-A2: 10%) forms. Then, the non-CLIL B1 and B2 groups show a different pattern: the non-CLIL-B1 produces NP (46.3%) and overt (45.9%) forms almost to the same extent and the non-CLIL-B2 group shows a native-like behaviour producing more NP (48%) forms, followed by overt (39.1%) and null (12.9%) forms. group shows a native-like behaviour producing more NP (48%) forms, followed by overt (39.1%) and null (12.9%) forms.

Using the English native speakers as a reference point, we can see that both CLIL and non-CLIL groups at B2 level behave in a native-like manner but importantly, the non-CLIL-B2 group shows a more native-like pattern. The other groups show variability and different patterns of production: the CLIL groups at A2 and B1 levels seem to opt for a higher production of overt forms, while the non-CLIL groups at A1 and A2 levels opt for a considerable higher production of NP forms.

CLIL	A1		A2		B1		B2		Native speakers	
Total frequencies	N=116		N=174		N=194		N=193		N=188	
	N	%	N	%	N	%	N	%	N	%
null	11	9.48%	19	10.92%	11	5.67%	10	5.18%	29	15.43%
overt	41	35.34%	86	49.43%	97	50.00%	85	44.04%	67	35.64%
NP	64	55.17%	69	39.66%	86	44.33%	98	50.78%	92	48.94%

Table 4. Overall use of REs across proficiency level in CLIL group

Non-CLIL	A1		A2		B1		B2		Native speakers	
Total frequencies	N=124		N=161		N=218		N=179		N=188	
	N	%	N	%	N	%	N	%	N	%
null	7	5.65%	16	9.94%	17	7.80%	23	12.85%	29	15.43%
overt	16	12.90%	49	30.43%	100	45.87%	70	39.11%	67	35.64%
NP	101	81.45%	96	59.63%	101	46.33%	86	48.04%	92	48.94%

Table 5. Overall use of REs across proficiency level in non-CLIL group

When we compare the production of REs in CLIL and non-CLIL groups, we can see some differences between them. In particular, A1 and A2 groups show significant differences in the use of overt and NP forms. The non-CLIL-A1 group shows a higher production of NP forms and a lower production of overt forms, which is statistically significant compared to the CLIL-A1 group. Also, the A2 groups show statistically significant differences for NP and overt forms as the non-CLIL-A2 group shows a higher production of NP forms, while the CLIL-A2 group shows a higher production of overt forms. These two groups of learners show differences between them; however, notice that they are far from native-like production and those differences are a reflection of their variability in the use of REs at initial stages but we cannot infer that one group performs better than the other showing an advantage due to longer exposure. When it comes to B1 and B2 groups, we can see that they do not show so many differences between them. In fact, it is the B2 level the one that shows differences in the production of null forms; importantly, the use of null forms in non-CLIL-B2 is higher (and it is statistically significant) compared to the CLIL-B2 group. Surprisingly, the non-CLIL-B2 group behaves more native-like than the CLIL-B2 group. Therefore, this only difference also shows that longer exposure does not seem to affect the use of REs.

6.3. RQ3. Production of REs According to Character

In order to address research question 3, this section shows the production of REs for the three most prominent characters in the story (i.e. the boy, the dog and the frog). If we first focus on native speaker production, Figure 3 (left and right charts) shows that the most produced form for 'the boy' is overt (43.5%), closely followed by NP (41.7%) and then null (14.8%). By contrast, Figure 4 and Figure 8 show that the preferred RE for 'the dog' and 'the frog' is NP, followed by overt and null. Section 6.2 above showed that the most produced form by native speakers was NP. Here, we can see that this is maintained for the dog and the frog, but there is a slight change for 'the boy', which to some extent shows that the type of character influences the choice of the REs.

Both CLIL and non-CLIL groups reveal a native-like tendency showing a higher production of NP forms for 'the dog' and 'the frog' but their production of NP forms to refer to these characters is even higher in some groups. When it comes to the boy, the CLIL group and the non-CLIL group at B1 and B2 levels show a higher production of overt forms, as native speakers do; by contrast, the non-CLIL at A1 and A2 levels evidence a preference for NP forms to refer to the boy. Apart from that difference, which may reflect an overproduction of NP forms at initial stages, both groups show a native-like tendency opting for a higher use of NP forms for the non-human characters and higher use of overt forms for the human character. Again, some differences between CLIL and non-CLIL groups emerge at initial levels but they perform similarly in the end. This shows that the longer exposure to the L2 does not affect the choice of the RE in discourse.
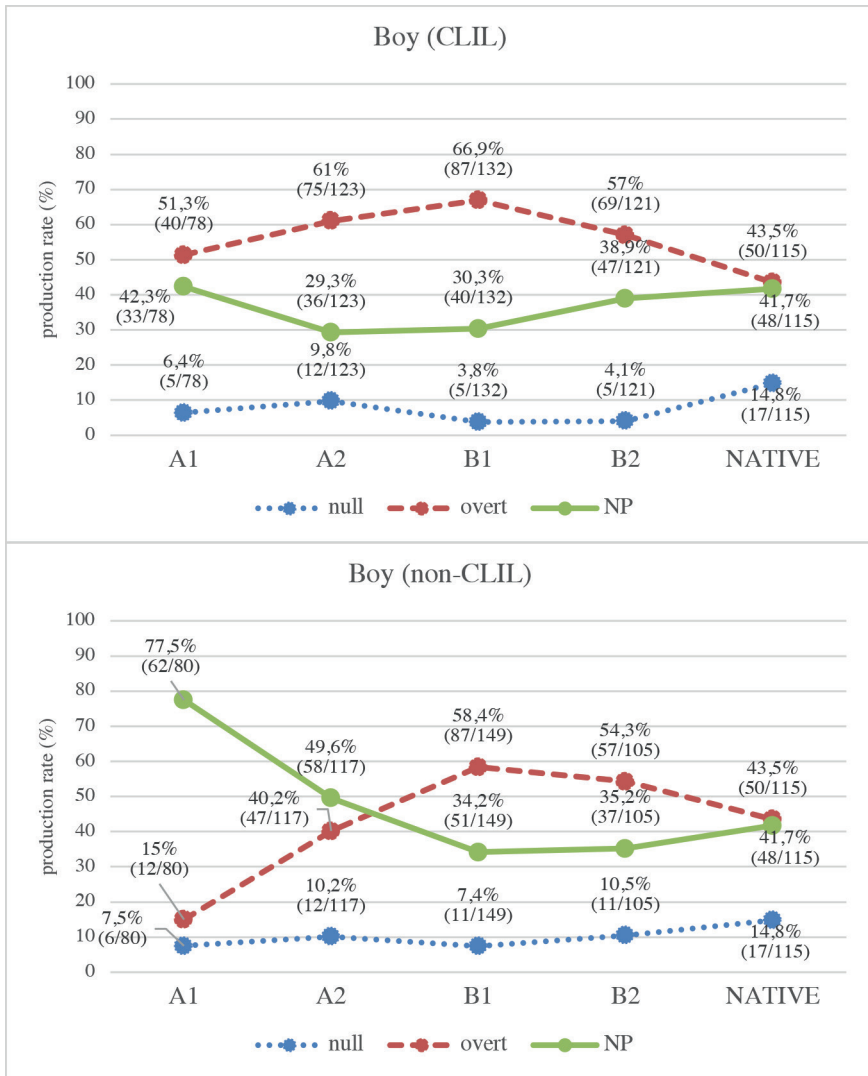
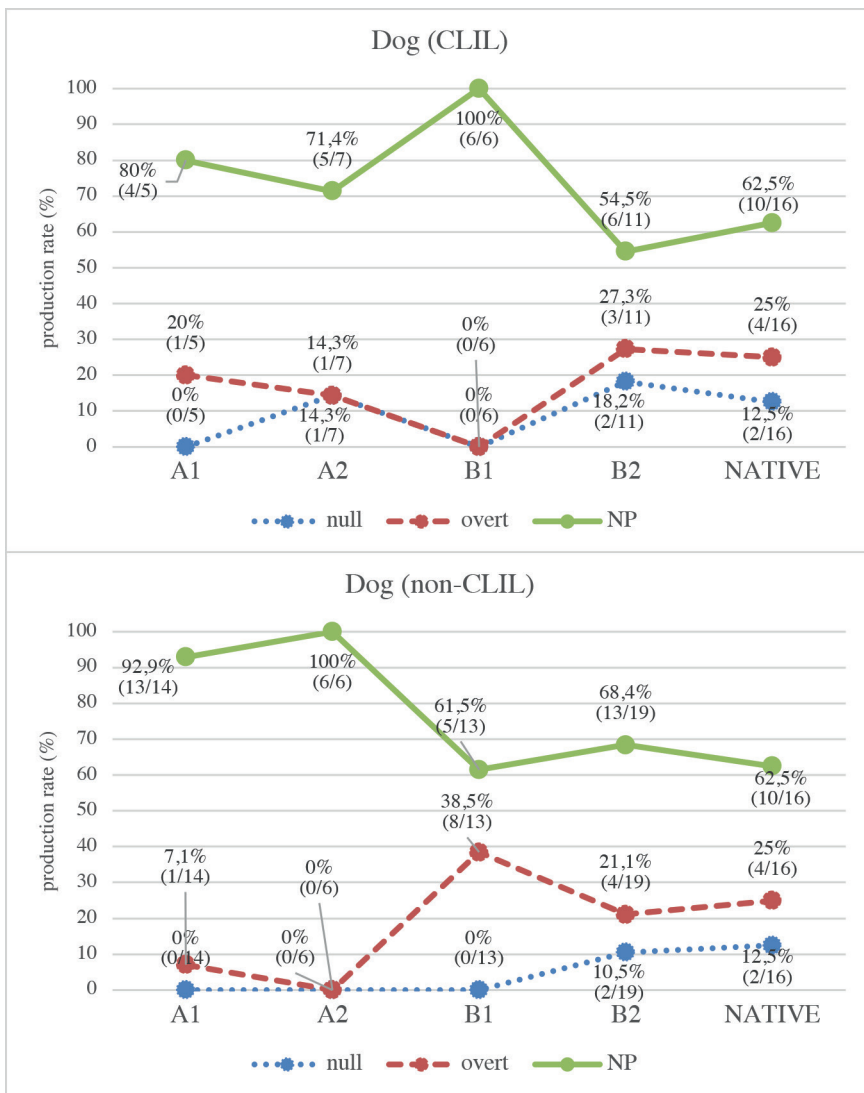Figure 3. Production of RE for boy in CLIL (first chart) and non-CLIL (second chart) groups

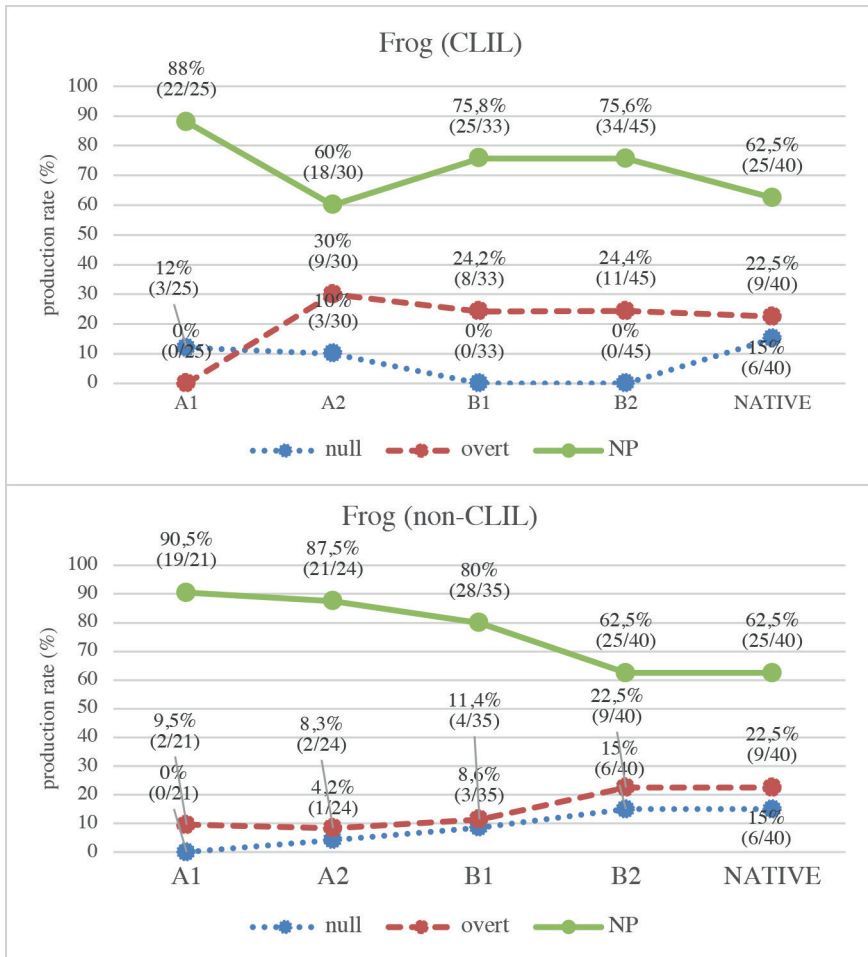Figure 4. Production of RE for dog in CLIL (first chart) and non-CLIL (second chart) groups

Figure 5. Production of RE for frog in CLIL (first chart) and non-CLIL (second chart) groups

6.4. RQ4. Production of RE According to Information Status

H4 predicted that in topic continuity contexts, less full forms would be used in order to achieve cohesion. If we first focus on native speaker production,

Figure 6 shows that in topic continuity contexts, native speakers produce more overt forms (53%), followed by null (29%) and NP (18%) forms. Therefore, we can confirm that they use more overt and null forms in topic continuity and therefore provide cohesion in their discourse. If we now turn to learner production, in topic continuity contexts, the CLIL group produces more overt forms (as native speakers do) followed by NP and null forms (opposite to what native speakers do except for the CLIL-A2 group that produces more null than NP forms). Notice that even though the CLIL group prefers overt forms in topic continuity contexts, their production is significantly higher than that of native speakers, which shows that the CLIL group is overusing these forms. In topic continuity contexts, the non-CLIL groups show a different pattern as Figure 7 shows. The non-CLIL-A1 group starts producing more NP forms followed by overt and null forms, which shows that they are overexplicit. The other non-CLIL groups produce more overt forms in topic continuity contexts (as native speakers do) and their production is just significantly higher for the non-CLIL-B1 group. Therefore, again we can observe an overuse of overt forms for the non-
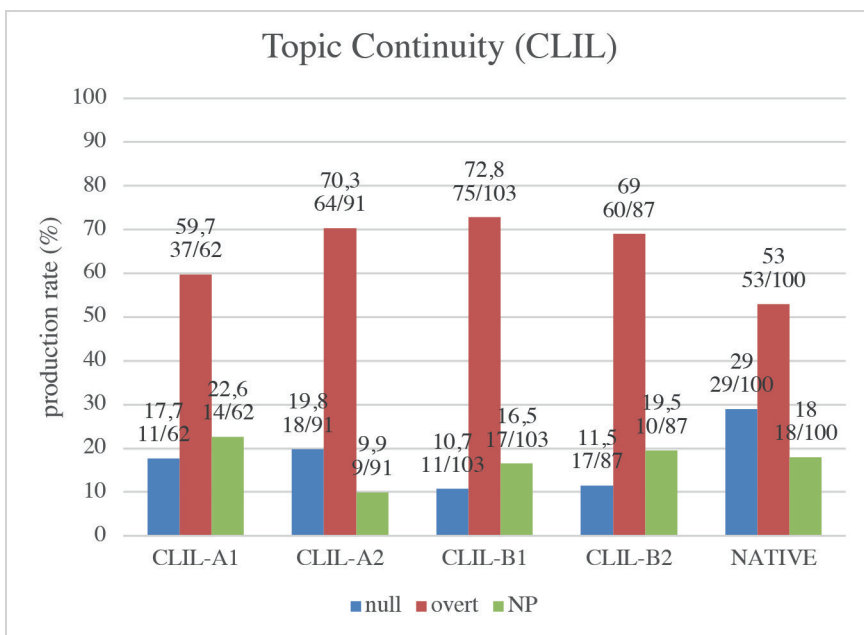

Figure 6. Production of RE in topic continuity for CLIL groups

CLIL-A1 and non-CLIL-B1. If we compare the production of CLIL and non-CLIL groups, we can see that there are significant differences between the CLIL and the non-CLIL groups at A1 and A2 levels in the production of overt and NP forms. The CLIL groups at these two levels behave in a more native-like manner and we could say that they outperform the non-CLIL groups. However, there are no significant differences between CLIL and non-CLIL groups in B1 and the differences found in B2 levels favour the non-CLIL group. The non-CLIL-B2 group produces more null forms following the native trend, while the CLIL-B2 produces more NP forms than overt. So, the non-CLIL groups does not only catch up the CLIL group at B1 and B2 levels, but even they show a more native-like behaviour.
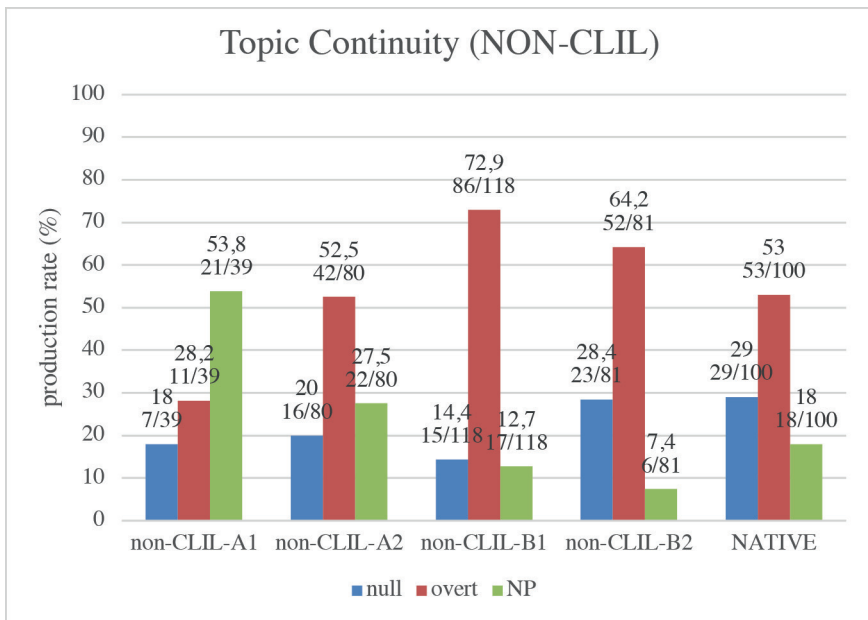

Figure 7. Production of RE in topic continuity for non-CLIL groups

7. Discussion

Our first research question was concerned with the word length and syntactic complexity of narratives. Our prediction was that learners would

differ from native speakers and that there would be differences between CLIL and non-CLIL groups at the same proficiency level, but none of these two predictions were confirmed. First, learners differ from native speakers in terms of length and syntactic complexity at beginner levels but not at later stages of proficiency. This shows that learners have developed their narrative abilities at this stage and it is possible to compare their production with native speaker production. Second, we expected CLIL learners to outperform non-CLIL learners but it seems that narratives are not affected by longer exposure to English throughout other curricular subjects as there are just some differences between CLIL and non-CLIL groups. In particular, the CLIL-A2 group produces more subordinate clauses than the non-CLIL-A2, and when comparing the two types of main sentences (main and coordinate), the CLIL-A1 group produces more compound sentences than their non-CLIL counterparts leading to a more fluid and cohesive discourse. So, we can say that the CLIL groups could be performing better at initial stages but then the non-CLIL groups catch up and they do not show differences. All in all, there seems to be no clear CLIL advantage for these results.

Research question 2 showed the overall production of REs. The predictions were that more overt and NP forms would be produced across all groups but there would be variability between learners and native speakers. Our predictions are partly confirmed as both native speakers and learners produce more overt and NP forms, which are the expected forms in a non-pro-drop language like English. However, different patterns of use are observed. Native speakers opt for a higher production of NP forms followed by overt and null forms. CLIL and non-CLIL learners at A1 and A2 levels show some differences: the CLIL-A1 group production differs from the non-CLIL-A1, being the former more native-like, and then the opposite occurs as the non-CLIL-A2 behaves like native speakers and its production is statistically significant compared to the CLIL-A2. Therefore, we can confirm that learners differ from native speakers but we cannot say that the CLIL group outperforms the non-CLIL group (or vice versa) as there is variability among them. This, thus, shows that both groups are still in the process of acquiring the use of REs. When it comes to CLIL and non-CLIL learners at B1 and B2 levels, we can see that all these groups follow native-like patterns and the only encountered difference between them favours the non-CLIL-B2 group. This group produces more null forms (in a native-like manner) and when it comes to statistics, their production is

significantly different compared to CLIL-B2, which shows a lower production of null forms. All in all, these results again show that CLIL instruction does not seem to yield benefits for the use of REs as both intermediate groups behave similarly and even it is the non-CLIL-B2 group the one that is closer to the native norm.

Research question 3 looked at the use of REs depending on the type of character. It was predicted that the type of character would determine the choice of the RE and also differences between native speakers and learners were expected. We first confirmed that the choice of the RE could be partly determined by the character as native speaker production shows: they produce more overt forms for the character 'the boy' and more NP forms for the characters 'the dog' and 'the frog'. Results in research question 2 showed that native speakers produced more NP forms overall but now there is a change in the most produced RE when we focus on different characters, which leads us to think that, to some extent, the character influences the choice of the RE. These results are in line with previous literature (Kang, 2004). CLIL and non-CLIL learners at B1 and B2 levels show the same pattern as native speakers, so their production is also influenced by the type of character and again, there are no differences between CLIL and non-CLIL so our prediction that they would differ is not confirmed. Importantly, CLIL and non-CLIL learners at A1 and A2 levels show differences among them. In this case, the CLIL groups at A1 and A2 levels show a native-like tendency but the non-CLIL groups at A1 and A2 levels opt for using more NP forms to refer to the character 'the boy'. This shows that they are overusing NP forms and it seems that now the CLIL group outperforms the non-CLIL group. At the end, however, both groups show a similar behaviour at intermediate level so the non-CLIL groups finally catch up the CLIL group.

Finally, our fourth research question looked at the use of REs in topic continuity contexts. We predicted that less full forms would be used by native speakers, and additionally, learners were predicted to differ from native speaker production. We confirmed that in topic continuity contexts, speakers achieve a cohesive discourse by producing less full forms and therefore native speakers produced a higher amount of overt and null forms in topic continuity contexts. Again, the use of the REs in this section differs from the results in section 6.2 so that we can see an additional factor that constrains the use of the RE. Learners show some differences compared to

native speakers in topic continuity contexts. In particular, CLIL learners at A1 and A2 levels initially seem to behave as native speakers (as they produce more overt forms), but their production of overt forms is significantly higher than that of native speakers. This shows that learners overproduce overt forms and they are in the process of acquiring a more pragmatic use of REs. Non-CLIL groups at A1 and A2 levels show more differences compared to native speaker production and thus the CLIL groups (at these proficiency levels) behave in a more native-like manner, which again seems to show that they perform better than the non-CLIL groups. However, CLIL and non-CLIL groups at B1 and B2 in topic continuity contexts again do not show differences and behave in a native-like manner. In fact, we can again notice that non-CLIL-B2 learners show higher production of null forms (as we showed above in 6.3) and behave more-native-like than the CLIL-B2 group. All in all, the CLIL groups at beginner levels seem to perform better than their non-CLIL counterparts; however, both CLIL and non-CLIL groups at intermediate levels behave similarly (and even the non-CLIL-B2 group is closer to the native norm), which again indicates that CLIL instruction does not pose benefits for the acquisition of REs.

8. Conclusion

This paper explored the written production of native English speakers and L2 English learners attending CLIL and non-CLIL programmes to ascertain possible differences in the overall production in narratives and the use of REs in discourse. Native English data was used as a point of reference in order to explore learner production; additionally, this study analysed two main groups of learners (CLIL and non-CLIL) at different proficiency levels in order to explore if length of exposure (through CLIL instruction) would affect the development of narratives in general and the use of REs in particular.

The results above showed that there are no significant differences between native speakers and learners (apart from those between CLIL and non-CLIL beginners) in terms of general features of narratives (i.e. word length and sentence type), which indicates that learners at B2 level have acquired general features of narratives. When it comes to investigating the use of REs in discourse, the results above confirmed that i) the factors

investigated (i.e. type of character and information status) affect the choice of REs in both native speakers and learners (although not always to the same extent) and ii) the production of REs between learners and native speakers and between CLIL and non-CLIL learners differ. In particular, learners tend to be over-specific in some contexts, though they show native-like patterns at intermediate proficiency levels, and CLIL and non-CLIL learners show differences between them, especially at beginner levels (A1 and A2). Research questions 3 and 4 showed that CLIL learners at A1 and A2 levels tend to outperform their non-CLIL counterparts; however, the non-CLIL groups catch up at B1 and B2 levels and behave as their CLIL counterparts and even the non-CLIL B2 group ends up behaving more native-like.

All in all, we can conclude that at initial stages CLIL instruction has a positive impact at the syntax-discourse interface that involves the use of REs. However, CLIL and non-CLIL learners handle the use of REs in a similar manner at intermediate level; that is, even though CLIL learners have had longer exposure to the L2, they end up performing same as their non-CLIL counterparts. Therefore, this type of exposure could be benefiting learners at beginner levels but at the end, non-CLIL learners reach the same performance with shorter exposure to the L2. Also, it is important to notice that intermediate groups (both CLIL and non-CLIL) behave similarly but they are still far from the native norm. This leads us to think that another type of input (or even instruction) could lead to an improvement in the use of REs and consequently a native-like performance. These assumptions are beyond the scope of this study and we leave them to be explored in further research.

Acknowledgments

References

Adrián, M. M., & Mangado, M. J. G. (2015). L1 Use, Lexical Richness, Accuracy and Syntactic Complexity in the Oral Production of CLIL and NON-CLIL Learners of English. *Atlantis. Journal of the Spanish Association for Anglo-American Studies*, *37*(2), 177-199–199.

Bruton, A. (2011a). Are the differences between CLIL and non-CLIL groups in Andalusia due to CLIL? A reply to Lorenzo, Casal and Moore (2010). *Applied Linguistics*, *32*(2), 236–241. https://doi.org/10.1093/applin/amr007

Bruton, A. (2011b). Is CLIL so beneficial, or just selective? Re-evaluating some of the research. *System*, *39*(4), 523–532. https://doi.org/10.1016/j. system.2011.08.002

Cambridge University Press. (2010). English unlimited placement test. Cambridge: Cambridge University Press.

Cenoz, J., Genesee, F., & Gorter, D. (2014). Critical Analysis of CLIL: Taking Stock and Looking Forward. *Applied Linguistics*, *35*(3), 243–262. https://doi.org/10.1093/applin/amt011

Crosthwaite, P. (2011). The Effect of Collaboration on the Cohesion and Coherence of L2 Narrative Discourse between English NS and Korean L2 English Users. *Asian EFL Journal*, *13*.

Crosthwaite, P. (2013). An error analysis of L2 English discourse reference through learner corpora analysis. *Linguistic Research*, *30*(2), 163–193.

García Mayo, M. del P., & Villarreal Olaizola, I. (2011). The development of suppletive and affixal tense and agreement morphemes in the L3 English of Basque-Spanish bilinguals. *Second Language Research*, *27*(1), 129–149. https://doi.org/10.1177/0267658310386523

Gelormini-Lezama, C., & Almor, A. (2011). Repeated Names, Overt Pronouns, and Null Pronouns in Spanish. *Language and Cognitive Processes*, *26*(3), 437–454. https://doi.org/10.1080/01690965.2010.495234

Gutiérrez-Mangado, M. J., & Martínez-Adrián, M. (2018). CLIL at the linguistic interfaces. *Journal of Immersion and Content-Based Language Education*, *6*(1), 85–112. https://doi.org/10.1075/jicb.17002.gut

Hendriks, H. (2003). Using nouns for reference maintenance: A seeming contradiction in L2 discourse. In A. Giacalone (Ed.), *Typology and Second Language Acquisition* (pp. 291–326). Berlin: Mouton De Gruyter.

Hendriks, P., Koster, C., & Hoeks, J. C. J. (2014). Referential choice across the lifespan: why children and elderly adults produce ambiguous pronouns. *Language and Cognitive Processes*, *29*(4), 391–407. https://doi.org/10.108 0/01690965.2013.766356

Hickmann, M., & Hendriks, H. (1999). Cohesion and anaphora in children's narratives: a comparison of English, French German, and Mandarin Chinese. *Journal of Child Language*, *26*(2), 419–452.

Jiménez Catalán, Ruiz de Zarobe, Y. R. de, & Cenoz, J. (2006). Vocabulary profiles in English as a foreign language at the end of Spanish primary and secondary education. *Vienna English Working Papers (VIEWZ) (Special Issue: Current Research on CLIL I)*, *15*(3).

Kang, J. Y. (2004). Telling a coherent story in a foreign language: analysis of Korean EFL learners' referential strategies in oral narrative discourse. *Journal of Pragmatics*, *36*(11), 1975–1990. https://doi.org/10.1016/j. pragma.2004.03.007

Lasagabaster, D. (2008). Foreign Language Competence in Content and Language Integrated Courses. *The Open Applied Linguistics Journal*, *1*(1).

Lázaro, A. (2012). Faster and Further Morphosyntactic Development of CLIL vs. EFL Basque-Spanish Bilinguals Learning English in High-School. *International Journal of English Studies*, *12*(1), 79–96. https://doi. org/10.6018/ijes.12.1.125621

Lázaro, A., & García, M. M. del P. (2012). L1 use and morphosyntactic development in the oral production of EFL learners in a CLIL context. *International Review of Applied Linguistics in Language Teaching*, *50*(2), 135–160. https://doi.org/10.1515/iral-2012-0006

Leclercq, P., & Lenart, E. (2013). Discourse Cohesion and Accessibility of Referents in Oral Narratives: A Comparison of L1 and L2 Acquisition of French and English. *Discours. Revue de Linguistique, Psycholinguistique et Informatique.*, (12). https://doi.org/10.4000/discours.8801

Llinares, A., & Whittaker, R. (2007). Linguistic analysis of secondary school students' oral and written production in CLIL contexts.

Lozano, C. (2009). Selective deficits at the syntax-discourse interface: evidence from the CEDEL2 corpus. In N. Snape, Y. I. Leung, & M. Sharwood Smith (Eds.), *Language Acquisition and Language Disorders* (Vol. 47, pp. 127–166). Amsterdam: John Benjamins Publishing Company.

Lozano, C. (2016). Pragmatic principles in anaphora resolution at the syntax-discourse interface: advanced English learners of Spanish in the CEDEL2 corpus.

Lozano, C., Díaz-Negrillo, A., & Callies, M. (forthcoming). Designing and compiling a learner corpus of written and spoken narratives: COREFL. In *C. Bongartz, & J. Torregrossa (Eds.), What's in a narrative? Variation in story-telling at the interface between language and literacy.* Frankfurt: Peter Lang.

Marsh, D. (2002). CLIL/EMILE - the European dimension : actions, trends and foresight potential.

Martínez Adrián, M. (2011). An overview of Content and Language Integrated Learning: origins, features and research outcomes.

Martínez Adrián, M., & Gutiérrez Mangado, M. J. (2015). Is CLIL instruction beneficial in terms of general proficiency and specific areas of grammar? *Journal of Immersion and Content-Based Language Education*, *3*(1), 51–76. https://doi.org/10.1075/jicb.3.1.03adr

Martínez-Adrián, M., & Gutiérrez-Mangado, M. J. (2009). The Acquisition of English syntax by CLIL learners. Content and Language Integrated Learning. In *Content and Language Integrated Learning: Evidence from Research in Europe* (pp. 176–196). Multilingual Matters.

Mayer, M. (1969). *Frog, where are you?* (Dial Books for Young Readers). New York.

O'Donnell, M. J. (2008). The UAM CorpusTool: Software for corpus annotation and exploration. *Proceedings of the XXVI Congreso de AESLA*, 1433–1447.

Pérez Cañado, M. L. P. (2018). CLIL and Educational Level: A Longitudinal Study on the Impact of CLIL on Language Outcomes. *Porta Linguarum: Revista Internacional de Didáctica de Las Lenguas Extranjeras*, (29), 51–70.

Pladevall Ballester, E. (2013). Adult instructed SLA of English subject properties.

Prentza, A. (2014). *Pronominal Subjects in English L2 Acquisition and in L1 Greek: Issues of Interpretation, Use and L1 Transfer*. (N. Lavidas, T. Alexiou, & A. M. Sougari, Eds.). London: Versita, Versita Ltd.

Ruiz de Zarobe, Y. R. de. (2011). Which language competencies benefit from CLIL? An insight into Applied Linguistics research | Request PDF. In *Content and Foreign Language Integrated Learning: Contributions to Multilingualism in European Contexts* (pp. 129–153). Peter Lang.

Ruiz de Zarobe, Y. R. de, & Jiménez Catalán, R. M.-A. J. (2009). *Content and Language Integrated Learning: Evidence from Research in Europe*. Multilingual Matters.

Ryan, J. (2015). Overexplicit Referent Tracking in L2 English: Strategy, Avoidance, or Myth? *Language Learning*, *65*(4), 824–859. https://doi.org/10.1111/lang.12139

Villarreal Olaizola, I., & García Mayo, M. del P. (2009). Tense and Agreement Morphology in the Interlanguage of Basque/Spanish Bilinguals: CLIL versus non-CLIL. In *Content and language integrated learning: evidence from research in Europe* (pp. 157–175). Multilingual Matters.

White, L. (2009). Grammatical theory: interfaces and L2 knowledge. In W. C. Ritchie & T. K. Bhatia (Eds.), *The New Handbook of Second Language Acquisition* (pp. 49–66). BRILL.

Wigglesworth, G. (1990). Children's narrative acquisition: a study of some aspects of reference and anaphora. *First Language*, *10*(29), 105–125. https://doi.org/10.1177/014272379001002902