The –*s* morpheme in CLIL and EFL: DP structure and time/untimed tasks

Jasna Toquero Unanue
Ekialdea State High School

María del Pilar García Mayo
Universidad del País Vasco (UPV/EHU)

Received: 2024-03-18 /Accepted: 2024-05-11
DOI: *https://doi.org/10.30827/portalin.vi43.30409*
ISSN paper edition: 1697-7467, ISSN digital edition: 2695-8244

ABSTRACT: This study investigated whether type of instruction (CLIL versus EFL) and different amount of exposure have any effect on the comprehension and written production of the English 3rd person singular –s morpheme. Furthermore, it examined whether separation of the head noun determining number and person agreement from Tense and error type (omission vs commission) might be variables that could affect the comprehension of that morpheme. Twenty-six Spanish EFL learners (age range 15-16) completed an Acceptability Judgment Task and a Fill-in the Gaps Task in a timed and untimed mode. Contrary to expectations, no significant differences were found between the two groups in the four tasks. Moreover, results seemed to suggest that EFL learners performed better overall than their counterparts. Finally, the hypothesis that the separation of the head noun from Tense could be a factor playing a role in the comprehension of the *–s* morpheme was partially confirmed.
Keywords: CLIL, EFL, -s morpheme, timed/untimed tasks, DP structure

El morfema *–s* en AICLE y ILE: La estructura de la DP y las tareas con/sin límite de tiempo

RESUMEN: Este trabajo investigó si el tipo de instrucción (AICLE versus ILE) y la cantidad de exposición a la lengua extranjera tienen algún efecto en la comprensión y la producción escrita del morfema de tercera persona de singular en inglés (-s). Además, también estudió si la separación del núcleo del sintagma nominal (que determina la concordancia de número y persona) de la categoría Tiempo y el tipo de error (omisión vs comisión) podrían ser variables que afecten la comprensión de dicho morfema. Veintiséis aprendices de ILE (entre 15-16 años) completaron una tarea de juicio de aceptabilidad y otra de rellenar el huecos con y sin límite de tiempo. A diferencia de lo que se esperaba, no hubo diferencias significativas entre los dos grupos en ninguna de las tareas. Además, los resultados parecen indicar que el alumnado ILE obtuvo resultados generales mejores que el alumnado AICLE. Finalmente, se confirmó parcialmente la hipótesis de que la separación entre el núcleo y Tiempo tiene importancia en la comprensión del morfema *–s*.
Palabras clave: AICLE, ILE, morfema –s, tareas con/sin tiempo, estructura de la FD

1. Introduction

Content and Language Integrated Learning (CLIL) is an educational approach that became one of the priorities of the European Commission for Education at the beginning

of the 21st century (Eurydice, 2006). It has been defined as a dual-focused "educational approach where curricular content is taught through the medium of a foreign language" (Dalton-Puffer, 2011, p. 183). CLIL rapidly moved into mainstream education (San Isidro & Calvo, 2012), however, as some researchers have warned, the rapid spread of CLIL has outpaced measures of its impact (Pérez-Cañado, 2012), so there is a danger that quality might be overtaken by quantity.

Different studies have explored CLIL's possibilities and constraints and the extent to which positive outcomes are due to more extensive exposure to the language. Although research has not reached definite conclusions about the specific gains students might achieve in this setting, on the whole, research outcomes have pointed towards beneficial effects of CLIL on overall language proficiency. Recent research in Spanish monolingual communities has shown that CLIL does not have a detrimental effect on L1 competence and does not weaken content learning (Pérez Cañado, 2023). Compared to other areas where CLIL and English as a Foreign Language (EFL) learners' production has been contrasted, little research has been done on morphosyntax (Basterrechea & García Mayo, 2014; García Mayo & Villarreal Olaizola, 2011), which seems to be one the areas not favourably affected (García Mayo & Villarreal Olaizola, 2011; Martínez Adrián & Gutiérrez Mangado, 2015). If one of the key features of CLIL instruction is its balance between content and language, that is, meaning and form, attention to both should be paid during lessons (Nikula, Dafouz, Moore & Smit, 2016). One of the few morphosyntactic features that has been explored in the literature is the third person singular morpheme –s. However, to the best of our knowledge, there is no study that has analyzed the comprehension of this morpheme in CLIL environments.

The present study investigated whether type of instruction (CLIL versus EFL), and different amount of exposure have any effect on the comprehension and written production of the English 3rd person singular *–s* morpheme. Furthermore, it examined if separation of the head noun determining number and person agreement from Tense might be a variable that could affect the comprehension of that morpheme.

2. Literature Review

2.1. The linguistic feature under study: English 3rd personal singular *-s*

Research in the field of second language acquisition (SLA) has shown that the third person singular *-s* morpheme is acquired late by L1 English learners (Brown, 1973), as well as by English L2 learners (Dulay & Burt, 1973). Moreover, even after years of exposure and explicit instruction, this morpheme has been shown to be problematic for Basque-Spanish EFL learners as well (García Mayo & Villarreal Olaizola, 2011; García Mayo, Lázaro Ibarrola & Liceras, 2005).

Several hypotheses have been advanced to explain why grammatical morphemes are acquired in a specific order such as the properties of the morphemes (Goldschneider & DeKeyser, 2001), the syntactic relationships in the sentence (Hawkins & Casillas, 2008), or even L1 influence (Luk & Shirai, 2009; Murakami & Alexopoulou, 2015). Goldschneider and DeKeyser (2001) proposed that the order of acquisition of grammatical morphemes is determined by the properties of the morphemes themselves (perceptual salience, semantic

complexity, morphophonological regularity, syntactic category and frequency) and by a combination of those properties. In their study, they reported that the 3rd person singular marker *–s* obtained the lowest score out of six morphological features in 11 out of the 12 morpheme studies reviewed. The 3rd person singular *–s* morpheme might be problematic because it is not perceptually salient and it frequently occurs in complex codas in final position of verbs in combination with other consonant sounds. Secondly, it is a semantically complex form, since it expresses person, number and tense. Thirdly, plural *–s*, possessive *–'s* and third person singular *–s* are all homophonous with each other. Furthermore, it is a functional category and a bound morpheme and, although it is a frequent morpheme, the link between form and meaning can be difficult to detect.

Another factor that could play a role in the rate of acquisition of some morphemes is the syntactic relationships in the sentence, specifically the distance between the head noun determining number and person agreement from Tense. Hawkins and Casillas (2008) tested this hypothesis with adult learners: 20 lower intermediate proficiency speakers of L2 English, 10 native speakers of Chinese, and 10 native speakers of Spanish, together with a control group of 10 English native speakers. The researchers reported the frequency in suppliance in speech of English copula and auxiliary *be* and affixal regular past *–ed* and 3rd person singular present tense *–s*. Hawkins and Casillas (2008) proposed that dependencies are formulated as context-sensitive statements about how a form is inserted into in the string of syntactic terminal nodes. Consider the following sentences:

(1) a) *My brother owns a house*
b) *The brother of my best friend owns a house*
c) *My best friend's brother owns a house*

In (1a) a simple Determiner Phrase (DP) is involved ('My brother'). In sentences (1b) and (1c) on the contrary, a complex DP is involved ('The brother of my best friend', 'My best friend's brother'). Hawkins and Casillas (2008) proposed that participants would supply the 3rd person singular *–s* morpheme more frequently in the former than in the latter. They hypothesized that separation of the head noun determining number and person agreement from Tense would lead either to a decrease in suppliance of the *–s* (omission) or to (mis-) agreement (commission) with the closest noun. The findings of the study showed that there was no decrease in the suppliance of the *–s* morpheme when there was a complex subject with a preceding genitive DP, and, as expected, there was a decrease in the suppliance of *-s* when a complex subject followed by a PP disrupted adjacency. Finally, results showed that disruption resulted in omission (e.g. '*My brother own a house') rather than in (mis-) agreement with the closest noun (e.g. '*The guests owns a house'). When *–s* was not supplied, a bare verb (e.g. 'own') was.

All in all, even though many hypotheses have been advanced to explain the morpheme order, '[i]n the end, there has been no simple satisfactory explanation for the sequence, and most researchers agree that the order is determined by an interaction among a number of different factors' (Lightbown & Spada, 2011, pp. 3-4).

2.2 Timed and untimed tasks to measure linguistic knowledge

As mentioned above, English learners have been claimed to have difficulty in acquiring the third person singular *–s* morpheme because they are unable to produce it accurately when communicating. However, that does not mean that they lack declarative knowledge of this feature. Ellis (2009a) developed a battery of tests that would provide relatively separate measures of implicit and explicit knowledge. A total of 111 participants (20 L1 English and 91 L2 English learners) took part in his study and were administered tests designed to provide measures of their knowledge of 17 grammatical structures, among them the third person singular *–s* morpheme. Five tasks, three timed (Elicited Oral Imitation Test, Oral Narrative Test, Grammaticality Judgment Test) and two untimed (Grammaticality Judgment Test, Metalinguistic Knowledge Test) were administered. The working hypothesis was that time-pressured tests would require learners to rely on their implicit knowledge because there was little opportunity to access metalinguistic knowledge. The unpressured tests were predicted to measure explicit knowledge because they involved a high degree of awareness and the participants would fall on their metalinguistic knowledge.

Results showed that learners' performance on the tests with and without time pressure differed, with participants performing significantly better on the untimed tests. A more detailed analysis of the Grammaticality Judgment Tests' (GJTs) scores indicated that they differed significantly on two dimensions, namely timed versus untimed and grammatical versus ungrammatical. In other words, participants performed significantly better not only in the Untimed Grammaticality Judgment Test (UGJT), but also in the grammatical sentences. In the case of the UGJT, in particular, the grammatical and ungrammatical sentences appeared to be measuring different constructs, with the latter providing a more convincing measure of explicit knowledge. Ellis (2009b) concluded that learning difficulty seemed to be different depending on which type of knowledge was involved. Regarding the third person singular morpheme, the findings pointed to the fact that it may be easy in terms of explicit knowledge but difficult in terms of implicit knowledge.

2.3 Research on morphosyntax in CLIL contexts

Very little research has been carried out on the impact of CLIL on morphosyntax (Dalton-Puffer, 2007 and Hüttner & Rieder-Bünemann, 2010, in Austria; Järvinen, 2010, in Finland). A number of studies have analysed if type of instruction affects the production of some grammatical morphemes, focusing mainly on the oral production of the third person singular *-s* morpheme. Thus, Martínez Adrián and Gutiérrez Mangado (2009) investigated the oral production of 9 CLIL and 10 EFL Basque-Spanish bilingual student and showed that CLIL learners significantly outperformed EFL learners only in the use of placeholders *is/he* (insertion of *is* or *he* after the subject and before the verb). That is, the CLIL group did not seem to need placeholders and displayed a more accurate agreement morphology. In a cross-sectional study, Villarreal Olaizola and García Mayo (2009) examined the acquisition of suppletive (auxiliary and copula *be*) and affixal forms (third person singular *-s* morpheme and past tense *–ed* morpheme) of 56 Basque-Spanish EFL learners in their last year of secondary education. They were asked to orally narrate Mayer's (1969) picture-story

'*Frog, where are you?'* individually. Results showed that the EFL group omitted the affixal forms *–s* and *–ed* significantly more frequently than the CLIL group and that, although affixal morpheme omission was high across the two groups, commission errors were almost non-existent. This led the authors to conclude that the interlanguage of the participants was not impaired at an abstract level but, rather, that it was a problem with realizing the morphological form of finite verbs.

In a longitudinal study, García Mayo and Villarreal Olaizola (2011) investigated the acquisition of the same morphological features by 78 CLIL and EFL Basque-Spanish learners. Contrary to the previous study, no statistically significant differences were found between the CLIL and EFL groups regarding the development of the suppletive and affixal forms.

Basterrechea and García Mayo (2014), on the other hand, analysed the noticing and written production of the 3rd person singular *–s* morpheme by 116 CLIL and EFL Basque-Spanish learners who were in their first year of post-compulsory secondary education. CLIL and EFL learners were asked to complete a dictogloss task in pairs and individually. The pairs interacted orally in dyads in order to complete a communicative task which had the morpheme *–s* as the syntactic target, although the learners were not aware of it. CLIL learners noticed and produced more instances of the 3rd person singular morpheme than their EFL counterparts but the differences were not statistically significant.

In conclusion, research so far seems to show that, although CLIL might be beneficial in some respects, its impact on morphoysntactic development is non-significant and thus further research is needed to provide more robust evidence in what sense or another.

3. The Present Study

3.1 Participants

Data from 26 Spanish EFL learners in their fourth year of compulsory high-school education (age range 15-16) were collected. They were divided into two groups on the basis of the educational approach they were enrolled in: CLIL (n=11) and mainstream EFL context (n=15). Both groups had begun learning English in primary school but they differed in the amount of hours of exposure to the foreign language. Thus, the EFL group had approximately 429 hours of exposure, with three hours of English instruction per week, whereas the CLIL had approximately 711 hours because they were taught Social Sciences in English. None of them was taking extramural English classes. The participants signed a consent form, were asked to provide general information in a background questionnaire and took the *Quick Oxford Placement Test* (OPT, Syndicate UCLE, 2001), which indicated that they had an A2 proficiency level.

3.2. Research questions

Two tasks, an acceptability judgment task (AJT) and a fill-in the gap task (FIG) task, in both timed (T) and untimed (U) versions, were administered. Based on the theoretical background and the findings from previous research, the following research questions were entertained:

1. Does separation of the head noun determining number and agreement from Tense lead either to a decrease in the identification of omission or commission errors?
2. Does error type (omission vs commission) have an effect on the accurate performance of the learners?
3. Does quantity of exposure (CLIL vs EFL) have an effect on learners' accurate performance in both the timed and untimed AJT and FGT?

3.3 Materials and procedure

The participants' performance was assessed on the basis of four pen-and-paper tasks completed individually: An AJT administered in a timed and an untimed version (TAJT, UAJT) and a FIG task also administered in two versions (TFGT, UFGT). The timed versions of both tasks were predicted to measure primarily implicit knowledge, whereas the untimed versions were predicted to measure explicit knowledge.

In the TAJT learners were allowed 20 minutes to indicate by ticking the relevant box whether each sentence was correct or incorrect and were asked to provide a correction for sentences rated as incorrect, to ensure that learners were rejecting ungrammatical sentences for the right reason (Ionin, 2012). Two types of fillers were included in the test: one that that looked similar on the surface but tested different phenomena, that is, the plural *-s* morpheme, and pronouns. The task consisted of 48 sentences evenly divided between grammatical (24) and ungrammatical (24). That is to say, for every ungrammatical sentence type tested, the corresponding grammatical sentence type was tested as well. There was also an equal number of experimental (24) and filler sentences (24). The latter included 12 grammatical and ungrammatical items featuring pronouns and 12 grammatical and ungrammatical sentences with the plural *–s* morpheme. There were 24 experimental sentences, 12 grammatical and 12 ungrammatical because the *-s* was missing in 6 of them (omission errors, e.g. '*My son's coach frighten him') or the *–s* was overused (commission errors, e.g. '*My girlfriend's cats hates me'). The participants were presented with a binary choice: Each item was scored dichotomously as correct/incorrect. Following Ionin (2012), no 'don't know' option was included in order to avoid learners overrelying on that option and to eliminate the negative feeling it produces not to know an answer. Sentences were equal in length (14 words) and all head nouns included in the preambles were animate. Collective nouns such as 'group', 'team', 'parents' were not included, given that subject-verb agreement with this type of nouns varies among speakers.

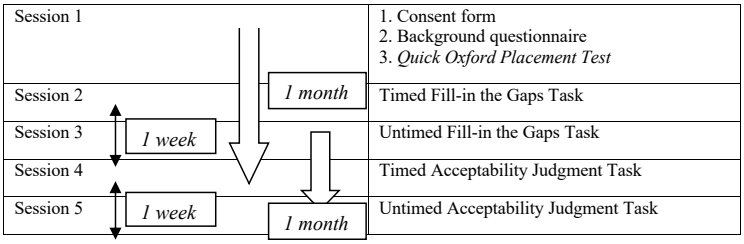
Based on Hawkins and Casillas (2008), sentences were divided into four categories: FILLERS-Simple DP ('She'); Complex DPs: DP's DP ('My cousins' neighbour), DP of DP ('The father of a friend') and FILLERS-Others ('My coach and my brother'). A total of 12 sentences (6 experimental and 6 filler sentences) belonged to each category. The UAJT contained the same 48 sentences as the TAJT, but learners were given as much time as they wanted to judge the correctness of sentences. To control for order of item presentation, two different, randomized versions of the same task were used.

The topic of the FGT was the difference between two twin sisters regarding their hobbies and the place they lived in. The text was adapted from Kenny and Luque-Mortimer (2004) and McKeegan (2004). The TFGT was a time-pressured task which allowed learners

15 minutes to provide the verbs between brackets in the correct tenses in the 48 gaps. In order to distract learners' attention from the target structure, a wide range of verb tenses was used. The experimental items were 16 sentences where the *–s* had to be provided in obligatory context. No auxiliaries or verbs in the negative or interrogative forms were categorized as experimental items. The UFGT contained the same 48 sentences and the same 16 experimental items as the TFGT.

In order to determine the length of time necessary to complete the tasks (20' for the TAJT and 15' for the TFGT), they were trialled on two English native speakers. The median response time was calculated for each task and an additional 20% was added to the median response time to each task to allow for the slower processing speed of L2 learners (Ellis 2009b). There was a one-week period between sessions to avoid learners relying on their working memory capacity because the interval between the tests cannot be too short if one wants to avoid producing a learning effect in the subjects, but it cannot be too long either because interlanguage grammars are in constant evolution (Sorace, 1996). Table 1 illustrates the procedure followed in the study:

Table 1. *Data collection procedure*



4. RESULTS

For the sake of clarity, data are presented in different sections, with each considering type of errors, correct suppliance of the *–s* and overall performance separately (significance set at p-value = .05. For research question 1 it was expected that both groups of learners would perform more accurately in the DP's DP category, as there is no intervening complement that disrupts adjacency. Table 2 features the scores of the DP's DP and DP of DP category for the CLIL and the EFL groups:

Table 2. *Number (#) and percentage (%) of errors committed when identifying and correcting omission and commission errors in the TAJT and the UAJT by the CLIL and EFL groups*

	CLIL								EFL							
	TAJT				UAJT				TAJT				UAJT			
	DP's DP		DP' of DP		DP's DP		DP' of DP		DP' of DP		DP's DP		DP's DP		DP' of DP	
	#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%
-s omission	22/22	100	22/22	100	22/22	100	21/22	95.45	27/30	90	27/30	90	25/30	83.33	25/30	83.33
-s commission	9/11	81.82	9/11	81.82	10/11	90.91	10/11	90.91	11/15	73.33	14/15	93.33	7/15	46.67	11/15	73.33

A one sample unilateral binomial test revealed that there were no significant differences between both categories in the CLIL group (TAJT: omission errors: z= 0, p-value= 0.50; commission errors: z= 0, p-value= 0.50. UAJT: omission errors: z= 1.0235, p-value= 0.85; commission errors: z= 0, p-value= 0.50). In the EFL group, the differences were significant both in the timed and untimed tasks only when identifying and correcting commission errors (TAJT: omission errors: z= 0, p-value= 0.50; commission errors: z= -3.1053, p-value= 0.0009. UAJT: omission errors: z=0, p-value= 0.50; commission errors: z= -2.3355, p-value= 0.01). This means that EFL learners identified and corrected commission errors significantly more frequently in the DP's DP category in the timed and untimed AJT. As can be seen in Table 2, the number of errors committed when identifying and correcting commission errors was 73.33% in the DP's DP category and 93.33% in the DP of DP category (TAJT) and 46.67% in the DP's DP category and 73.33% in the DP of DP category (UAJT). Nevertheless, care should be taken when analyzing the results as sample size may be too small for normal approximation.

In order to determine whether the two groups differed significantly on the performance in the DP's DP category versus DP of DP category, a two independent sample unilateral binomial test was used to test their performance when faced with omission and commission errors in each category. Contrary to expectations, no significant differences were found between the groups either in the timed or untimed AJT (TAJT, DP's DP: omission errors: z= -1.53, p-value= 0.06; commission errors: z= 0.51, p-value= 0.69. DP of DP: omission errors: z= 1.53, p-value= 0.94; commission errors: z= -0.91, p-value= 0.18. UAJT, DP's DP: omission errors: z= 1.85, p-value= 0.97; commission errors: z= 2.01, p-value= 0.98. DP of DP: omission errors: z= 2.34, p-value= 0.99; commission errors: z= 1.35, p-value= 0.91).

Moreover, results seem to suggest that the reverse may be the case, as EFL learners' performance was more accurate than the CLIL group performance in all the categories except in the TAJT, in the DP of DP category, when faced with commission errors: CLIL learners did not identify and correct 81.82% of the commission errors, whereas the EFL group did not identify and correct 93.3% of the commission errors.

The second research question asked whether error type would have an effect on the learners' accurate performance. A one sample unilateral binomial test showed that differences were significant in the CLIL and EFL group, both in the timed and untimed AJT. That is to say, both groups identified and corrected commission errors significantly more frequently than omission errors (CLIL group, TAJT: z= -4.4059, p-value= <0.0001. UAJT: z= -5.2855, p-value= <0.0001. EFL group, TAJT: z= -4.1404, p-value= <0.0001. UAJT: z= -6.1274, p-value= <0.0001). Table 3 illustrates the accurate responses when identifying omission and commission errors in the TAJT and UAJT by CLIL and EFL learners.

Table 3. *Number (#) and percentage (%) of accurate responses when identifying and correcting omission and commission errors in the TAJT and the UAJT by CLIL and EFL learner*

	CLIL								EFL							
	TAJT				UAJT				TAJT				UAJT			
	Omission		Commission		Omission		Commission		Omission		Commission		Omission		Commission	
	#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%
-s	0/66	0	15/66	22.72	1/66	1.52	21/66	31.81	9/90	10	27/90	30	13/90	14.44	42/90	46.67

In order to determine whether the two groups differed significantly on the performance when faced with omission and commission errors, a two independent sample unilateral binomial test was used. Contrary to expectations, no significant differences were found between the groups in the timed or untimed AJT (TAJT, omission errors: z= -2.65, p-value= 0.99; commission errors: z= -1.01, p-value= 0.84. UAJT, omission errors: z= -2.79, p-value= 0.99; commission errors: z= -1.86, p-value= 0.97). Again results seem to suggest that the reverse may be the case, as EFL learners' performance was more accurate than their CLIL counterpart's performance when faced with omission and commission errors both in the timed and untimed AJT.

We also wondered whether quantity of exposure would have any effect on learners' accurate performance in both the timed and untimed AJT. Contrary to expectations, a two independent sample unilateral binomial test revealed that there were no significant differences between the CLIL and the EFL group in any of the categories compared in the timed and untimed AJT: correct comprehension of the *–s* morpheme (TAJT: z= -0.8778, p-value= 0.81. UAJT: z= -1.3846, p-value= 0.92), performance when faced with omission errors (TAJT: z= -2.6465, p-value= 0.99. UAJT: z= -2.791, p-value= 0.99), performance when faced with commission errors (TAJT: z= 1.0118, p-value= 0.84. UAJT: z= 1.8674, p-value= 0.97) and overall performance (TAJT: z= -1.6682, p-value= 0.95. UAJT: z=-3.0220, p-value= 0.99). Figures 1, 2, 3 and 4 feature the mean average results by CLIL and EFL learners in the TAJT and UAJT.

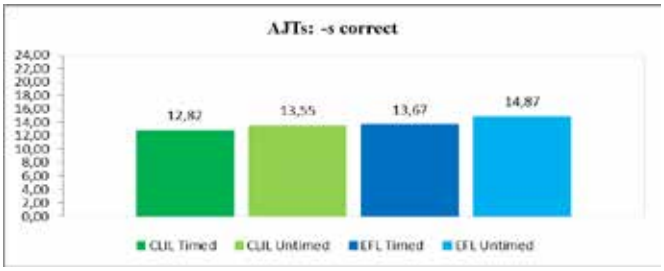


Figure 1. *Mean average results of the correct comprehension of the –s morpheme in the TAJT and UAJT by CLIL and EFL learners*

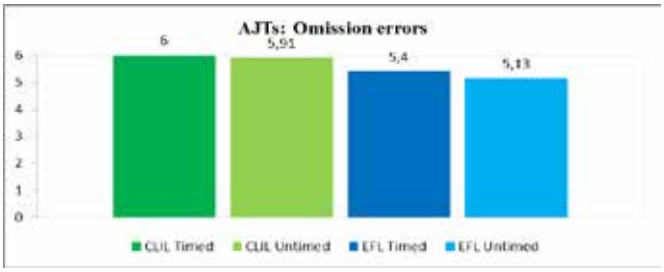


Figure 2. *Mean average results of errors comitted when faced with omission errors in the TAJT and UAJT by CLIL and EFL learners*

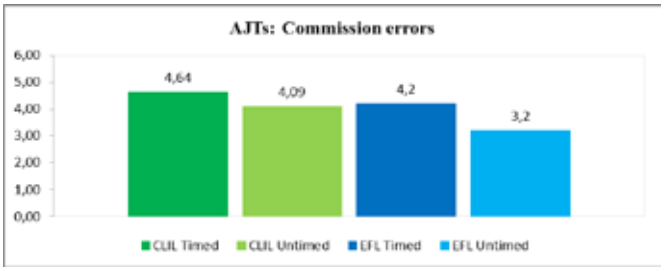


Figure 3. *Mean average results of errors comitted when faced with commission errors in the TAJT and UAJT by CLIL and EFL learners*

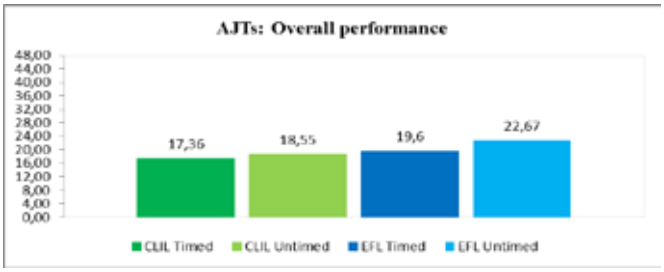


Figure 4. *Mean average results of the correct overall performance in the TAJT and UAJT by CLIL and EFL learners*

Not only were there no significant differences found in any of the categories between CLIL and EFL learners, but results seemed to suggest that EFL learners performed more accurately than their counterparts in all the categories. For example, Table 4 shows the errors committed by CLIL and EFL groups when faced with omission errors. Whereas the CLIL group neither identified nor corrected any of the omission errors in the TAJT, the EFL group identified 9 omission errors. In the UAJT, the CLIL group only identified one omission error, whereas the EFL group identified 13 omission errors.

Table 4. *Number (#) and percentage (%) of errors committed when faced with omission errors in the TAJT and UAJT by CLIL and EFL learners.*

	TAJT				UAJT			
	CLIL		EFL		CLIL		EFL	
	#	%	#	%	#	%	#	%
-s omission	66/66	100	81/90	90	65/66	98.48	77/90	85.56

The third research question asked whether quantity of exposure would have any effect on learners' accurate performance in both the timed and untimed FGT. A two independent sample unilateral binomial test showed no significant differences between the CLIL and the EFL group in any of the categories compared in the timed and untimed FGT: correct inflection of the *–s* morpheme (TFGT: z= -4.0768, p-value= 0.99. UFGT: z= -1.1980, p-value=

0.88), omission errors (TFGT: z= -5.7577, p-value= 0.99. UFGT: z= -2.3639, p-value= 0.99) and overall performance (TFGT: z= -3.2174, p-value= 0.99. UFGT: z= -1.9082, p-value= 0.97). Figures 5, 6 and 7 feature the mean average results by CLIL and EFL learners in the TFGT and UFGT.

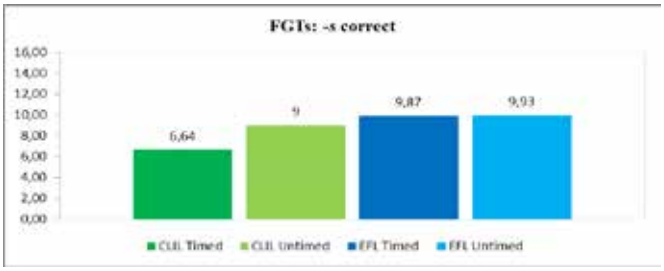


Figure 5. *Mean average results of the correct suppliance of the –s morpheme in obligatory contexts in the TFGT and UFGT by CLIL and EFL learners*

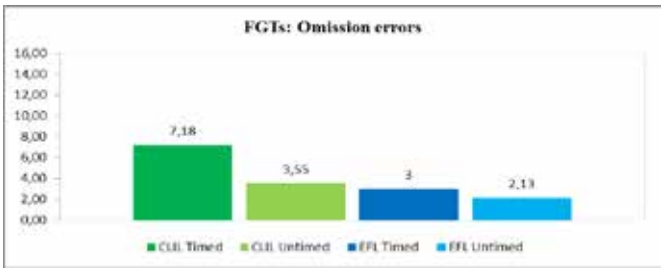


Figure 6. *Mean average results of omission errors comitted in the TFGT and UFGT by CLIL and EFL learners*

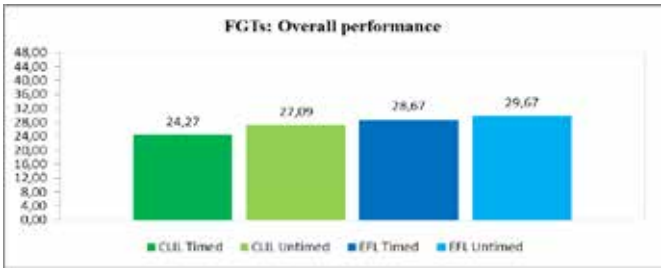


Figure 7. *Mean average results of the correct overall performance in the TFGT and UFGT by CLIL and EFL learners*

Our findings seem to indicate that EFL learners performed more accurately than their counterparts in all categories. Based on previous research (Ellis, 2009b), the present study tried to examine if time pressure and task stimulus had any effect on our participants'

performance. Regarding task stimulus, a one sample unilateral binomial test showed that differences were significant in the CLIL and EFL group, both in the timed and untimed AJT. That is to say, both groups performed significantly more accurately on the grammatical than on the ungrammatical sentences (CLIL group, TAJT: z= 30.4418, p-value= <0.0001. UAJT: z= 24.5227, p-value= <0.0001. EFL group, TAJT: z= 24.7831, p-value= <0.0001. UAJT: z= 18.2443, p-value= <0.0001). Table 5 illustrates the accurate responses when faced with grammatical and ungrammatical sentences in the TAJT and UAJT by CLIL and EFL learners.

Table 5. *Number (#) and percentage (%) of correct responses when faced with grammatical vs ungrammatical sentences in the TAJT and the UAJT by the CLIL and the EFL groups*

	CLIL								EFL							
	TAJT				UAJT				TAJT				UAJT			
	Grammatical		Ungrammatical		Grammatical		Ungrammatical		Grammatical		Ungrammatical		Grammatical		Ungrammatical	
	#	%	#	%	#	%	#	%	#	%	#	%	#	%	#	%
-s	126/132	95. 45	15/132	11. 36	127/132	96. 21	22/132	16. 67	169/180	93. 89	36/180	20	168/180	93. 33	55/180	30. 56

The CLIL group performed better in the grammatical sentences in both the TAJT and UAJT (TAJT: 95.45% correct performance the CLIL group and 93.89% correct performance the EFL group. UAJT: 96.21% correct performance the CLIL group and 93.33% correct performance the EFL group). The EFL group, on the other hand, performed better in the ungrammatical sentences both in the TAJT and UAJT (TAJT: 11.36% correct performance the CLIL group and 20% correct performance the EFL group. UAJT: 16.67% correct performance the CLIL group and 30.56% correct performance the EFL group).

Regarding time pressure, significant differences were partially found. A one sample unilateral binomial test showed that CLIL learners performed significantly better only in the UFGT, both in the experimental items (AJT: z= -0.9930, p-value= 0.16. FGT: z= -3.9506, p-value= 0.0001) and in the task as a whole (AJT: z= -1.1629, p-value= 0.12. FGT: z= -2.7209, p-value= 0.0033). Table 6 features feature the scores of the correct performance in the experimental items and in the whole tasks for the timed and untimed AJT and FGT for the CLIL group.

Table 6. *Number (#) and percentage (%) of correct performance in the –s morpheme and overall performance by the CLIL group in the timed and untimed AJT and FGT*

	CLIL							
	TAJT				UAJT			
	Timed		Untimed		Timed		Untimed	
	#	%	#	%	#	%	#	%
-s	141/264	53.41%	149/264	56.44%	73/176	41.48%	99/176	56.25%
Overall performance	191/528	36.17%	204/528	38.64%	267/528	50.57%	298/528	56.44%

The EFL group, on the contrary, performed significantly better only in the UAJT, both in the experimental items (AJT: z= -1.9539, p-value= 0.02. FGT: z= -0.1330, p-value= 0.44) and in the task as a whole (AJT: z= -3.4339, p-value= 0.0003. FGT: z= -1.1506, p-value= 0.12). Tables 7 features the scores of the correct performance in the experimental items and in the whole tasks for the timed and untimed AJT and FGT for the EFL group:

Table 7. *Number (#) and percentage (%) of correct performance in the –s morpheme and overall performance by the EFL group in the timed and untimed AJT and FGT*

	CLIL							
	TAJT				UAJT			
	Timed		Untimed		Timed		Untimed	
	#	%	#	%	#	%	#	%
-s	205/360	56.94%	223/360	61.94%	148/240	61.67%	149/240	62.08%
Overall performance	294/720	40.83%	340/720	47.22%	430/720	59.72%	445/720	61.81%

The tables above also show that the EFL group performed better than the CLIL group both in the timed and untimed AJT and FGT in the experimental items and in the task as a whole.

5. Discussion

The main goal of the present study was to compare the comprehension and written production of the third person singular *–s* morpheme by EFL and CLIL learners. Additionally, the study was also set out to test Hawkins and Casillas' (2008) findings. Our results could partially confirm Hawkins and Casillas' (2008) claims, as there were better results when adjacency was not disrupted. We need to consider, though, that theirs was an oral task and their participants were adult learners (age range 22-25). Regarding the comparison between CLIL and EFL learners, this study could not report significant differences between the two groups in their performance in the DP's DP category versus the DP of DP category.

As for the second research question, in line with previous studies (Hawkins & Casillas, 2008; Villarreal Olaizola & García Mayo, 2009) results showed that both groups of learners performed significantly better when identifying and correcting commission errors in the TAJT and UAJT. These results bear out the hypothesis that error type has an effect on learners' accurate performance. Comparing CLIL and EFL learners' performance when faced with omission and commission errors in the TAJT and UAJT, no significant differences were found between the two groups. We had also predicted that CLIL learners would perform more accurately than EFL learners in the timed and untimed AJT and FGT but, contrary to expectations, the scores in both groups consistently showed that there were no significant differences between the groups in the four tasks. In fact, the reverse seems to be the case, as EFL learners obtained better scores than CLIL learners in all the categories compared.

All in all, the results suggest that approximately 282 hours of extra exposure to the foreign language in the CLIL group do not seem to alter the comprehension and written production of the third person singular *–s* morpheme, since no significant differences were found in any of the four tasks between the two groups.

Coyle (2007) warned that, due to CLIL's rapid expansion, quality might have been overtaken by quantity in our particular setting. That is to say, the CLIL educational approach might have been understood as the teaching of non-language subjects in an additional language in the same way as the mother tongue, even though 'CLIL is not simply education in an additional language, it is education through an additional language based on connected

pedagogies and using contextual methodologies' (Coyle et al. 2010: 12). Nevertheless, these findings may also suggest that the differences reported in the studies comparing CLIL and EFL learners which indicated an advantage for the former might in fact be attributable to CLIL itself or to other variables such as the methodology followed or teacher training, rather than to the difference in the amount of hours of exposure. It should be remembered that, in Spain, CLIL teachers are usually non-native speakers of the target language and they are subject specialists rather than language specialists, which entails that they often have limited experience with matters of language proficiency and its development. Therefore, little focus on form (FonF) (Long, 1996) is found in the input teachers address to their learners, even though it has been proposed that explicit attention to language is of benefit to content learning (Alonso Pena & Pladevall-Ballester, 2020). Thus, many researchers advocate for the incorporation of FonF (Lyster, 2007) in CLIL programmes. The results obtained in this study also seem to point to the need for a more FonF approach in CLIL programmes.

Finally, this study considered the effect of time pressure and task stimulus on the learners' accurate performance. The findings showed that both variables are important because both groups of learners responded differently to grammatical and ungrammatical sentences in the TAJT and UAJT. CLIL and EFL learners performed significantly better in the grammatical sentences, that is, the task stimulus had an effect on learners' accurate performance. In the case of time pressure, the CLIL group performed significantly better in the UFGT, and no significant differences were found between the TAJT and the UAJT. The EFL group, on the contrary, performed significantly better in the UAJT, and no significant differences were found between the TFGT and UFGT. These findings are partially in line with Ellis (2009b), as both groups performed significantly better in the untimed mode but each of the groups in a different task: CLIL in the FGT and EFL in the AJT

This study has limitations that need to be acknowledged. Firstly, it was a cross-sectional study with a small sample size and a particular age group. Longitudinal studies with larger populations of different age groups should be carried out to validate these findings. Moreover, it would be convenient to have groups with higher proficiency level in the foreign language and to test if those learners would have fewer difficulties in the comprehension and written production of the *–s* morpheme. Furthermore, other grammatical morphemes have been included.

In view of the results, it is obvious that further research is needed to reach more definite conclusions on the effects of CLIL instruction on morphosyntax. Further research should explore the impact of task type and collect other types of data that measure both implicit and explicit knowledge in a timed and untimed mode. Besides, the use of on-line comprehension tasks as opposed to off-line tasks would be of benefit, as they are implicit tasks relatively immune to metalinguistic knowledge and response strategies and thus, they can better reflect the underlying process involved when participants process language in real-time.

6. References

Alonso Pena, C., & Pladevall-Ballester, E. (2020). Effects of focus on form on primary CLIL students' foreign language performance in task-based oral interaction. *Journal of Immersion and Content-based Language Education, 8,* 53-79. https://doi.org/10.1075/jicb.17028.alf

Basterreceha, M., & García Mayo, M.P. (2014). Dictogloss and the production of the English third person –s by CLIL and mainstream learners: A comparative study. *International Journal of English Studies, 14(2)*, 77-98. https://doi.org/10.6018/j.177321

Brown, R. 1973. *A first language: The early stages*. George Allen & Unwin Ltd.

Coyle, D. (2007). The CLIL quality challenge. In D. Marsh, & D. Wolff (Eds.), *Diverse contexts-converging goals. CLIL in Europe* (pp. 47-58). Peter Lang.

Coyle, D., Hood, P., & Marsh, D. (2010). *CLIL: Content and language integrated learning.* Cambridge University Press.

Dalton-Puffer, C. (2007). *Discourse in content and language integrated learning (CLIL) classrooms*. John Benjamins.

Dalton-Puffer, C. (2011). Foreword. In Y. Ruiz de Zarobe, J.M. Sierra, & F. Gallardo del Puerto (Eds.), *Content and foreign language integrated learning: Contributions to multilingualism in European contexts* (pp. 9-10). Peter Lang.

Dulay, H. C., & M. K. Burt. (1973). Should we teach children syntax? *Language Learning, 23*, 245-258. https://doi.org/10.1111/j.1467-1770.1973.tb00659.x

Ellis, R. (2009a). Implicit and explicit learning, knowledge and instruction. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders. *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 3-25). Multilingual Matters.

Ellis, R. (2009b). Measuring implicit and explicit knowledge of a second language. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders. *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 31-64). Multilingual Matters.

Eurydice. (2006). *Content and language integrated learning: Evidence from research in Europe.* Eurydice European Unit.

García Mayo, M.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/10.1177/026765831038652

García Mayo, M.P. (2005). Place holders in the English interlanguage of bilingual (Basque/Spanish) children. *Language Learning 55(3),*445-489. https://doi.org/10.1111/j.0023-8333.2005.00312.x

Goldschneider, J. M., & R. M. DeKeyser. (2001). Explaining the "natural order of L2 morpheme acquisition" in English: A meta-analysis of multiple determinants. *Language Learning, 55*, 1-50. https://doi.org/10.1111/1467-9922.00147

Hawkins, R., & G. Casillas. (2008). Explaining frequency of verb morphology in early L2 speech. *Lingua, 118,* 595-612.

Hüttner, J., & Rieder-Bünemann, A. (2010). A cross-sectional analysis of oral narratives by children with CLIL and non-CLIL instruction. In C. Dalton-Puffer, T. Nikula, & U. Smit (Eds.), *Language use and language learning in CLIL classrooms* (pp. 61-104). John Benjamins.

Ionin, T. (2012). Formal theory-based methodologies. In A. Mackey, & S. M. Gass (Eds.), *Research methods in second language acquisition: A practical guide* (pp. 30-52). Wiley-Blackwell.

Järvinen, H. M. (2010). Language as a meaning making resource in learning and teaching content: Analysing historical writing in content and language integrated learning. In C. Dalton-Puffer, T. Nikula, & U. Smit (Eds.), *Language use and language learning in CLIL classrooms* (pp. 145-168). John Benjamins.

Kenny, N., & Luque-Mortimer, L. (2004). *First certificate practice tests: Plus 1*. Pearson Education Limited.

Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie, & T. K. Bhatia (Eds.). *Handbook of second language acquisition* (pp. 413-468). Academic Press.

Lightbown, P. & Spada, N. (2011). *How languages are learned*. Oxford University Press.

Luk, Z.P., & Shirai, Y. (2009). Is the acquisition order of grammatical morphemes impervious to L1 knowledge? Evidence from the acquisition of plural *–s*, articles, and possessive *'s*. *Language Learning 59,* 721-754 10.1111/j.1467-9922.2009.00524.x

Lyster, R. (2007). *Learning and teaching languages through content. A counterbalanced approach*. John Benjamins.

Martínez Adrián, M., & Gutiérrez Mangado, M.J. (2009). The acquisition of English syntax by CLIL learners in the Basque Country. In Y. Ruiz de Zarobe, & R. Jiménez Catalán (Eds.) *Content and language integrated learning: Evidence from research in Europe* (pp. 176-196). Multilingual Matters.

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

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

McKeegan, D. (2004). *First certificate Passkey. Grammar section*. Macmillan Heinemann ELT.

Murakami, A., & Alexopoulou, T. (2015). L1 influence on the acquisition order of English grammatical morphemes. *Studies in Second Language Acquisition, 38,* 365-401. https://doi.org/10.1017/S0272263115000352

Nikula, T., Dafouz, E., Moore, P., & Smit, U. (Eds.). (2016). *Conceptualizing integration in CLIL and multilingual education*. Multilingual Matters.

Pérez Cañado, M. L. (2012). CLIL research in Europe: Past, present and future. *International Journal of Bilingual Education and Bilingualism, 15,* 315-341. https://doi.org/10.1080/13670050.2011.630064

Pérez Cañado, M.L. (2023). Inclusion and diversity in bilingual education. A European comparative study. *International Journal of Bilingualism and Bilingualism, 26,* 1129-1145 https://doi.org/10.1080/13670050.2021.2013770

San Isidro, X., & Calvo, E. (2012). The fusion effect of CLIL on language-building and content-learning: A research study. In D. Marsh, & O. Meyer (Eds.) *Quality interfaces: Examining evidence & exploring solutions in CLIL* (pp. 103-140). Eichstaett Academic Press.

Sorace, A. (1996). The use of acceptability judgments in second language acquisition research. In W. C. Ritchie, & T. K. Bhatia (Eds.). *Handbook of second language acquisition* (pp. 375-409). Academic Press.

Sydicate, U.C.L.E. (2001). *Quick placement test*. Oxford University Press.

Villarreal Olaizola, I., & García Mayo, M.P. (2009). Tense and agreement morphology in the interlanguage of Basque/Spanish bilinguals: CLIL vs. non-CLIL.In Y. Ruiz Zarobe & R. Jiménez Catalán (Eds.). *Content and language integrated learning: Evidence from research in Europe* (pp. 157-175). Multilingual Matters.