

A comparative study of language complexity and CDFs between students from areas with different socio-economic status and bilingual modalities

Marta González Romero

Máster en Lingüística Aplicada al Inglés



**MÁSTERES
DE LA UAM
2021-2022**

Facultad de Filosofía y Letras



**A comparative study of language complexity and
CDFs between students from areas with different
socio-economic status and bilingual modalities**

Marta González Romero

Tutora: Dra. Ana Llinares García

Master' s Degree in English Applied Linguistics

Facultad de Filosofía y Letras

Universidad Autónoma de Madrid

Academic year 2021/2022

Abstract

In recent decades, CLIL programmes have elicited a growing interest for bringing bilingual education possible and accessible to every student in a multicultural society. One of the main challenges that this approach face is the necessity to provide students with linguistic resources to express specific content on their L2. It is also necessary to take into account the characteristics of different CLIL programmes. Therefore, the present study aims to compare lexical density, *academicness* and grammar complexity between two bilingual modalities in the Community of Madrid bilingual programme: Bilingual Section and Bilingual Programme. Moreover, CLIL has been observed as the solution to eliminate the “discriminatory effect” in which socioeconomic status correlates with academic proficiency. Therefore, lexical density, *academicness* and grammar complexity are also analysed in two bilingual schools that differ in their socioeconomic status, later referred as Madrid West and Madrid South secondary schools. The results of the research show differences and similarities across bilingual modalities and secondary schools. The results lead to pedagogical and empirical implications and propose avenues for further research.

Keywords: Content and Language Integrated Learning (CLIL), lexical density, *academicness*, grammatical complexity, Cognitive Discourse Functions (CDFs), Systemic Functional Linguistics (SFL)

Resumen

En las últimas décadas, los programas AICLE han suscitado un creciente interés por hacer posible y accesible la educación bilingüe a todos en una sociedad multicultural. Sin embargo, uno de los principales desafíos a los que se enfrenta este enfoque es la necesidad de proporcionar recursos lingüísticos a los alumnos para expresar contenidos específicos en su L2. Además, es necesario tener en cuenta las características de los diferentes programas AICLE. Por ello, el presente estudio pretende comparar la densidad léxica, la *academividad* y la complejidad gramatical de dos modalidades bilingües del programa bilingüe de la Comunidad de Madrid: Sección Bilingüe y Programa Bilingüe. Además, se ha observado que los programas AICLE son clave para la eliminación del "efecto discriminatorio", en el que el estatus socioeconómico está correlacionado con la competencia académica. Por lo tanto, la densidad léxica, la *academividad* y la complejidad gramatical también se analizan dentro de dos contextos que difieren en su estatus socioeconómico, posteriormente referidos como institutos de secundaria Madrid Oeste y Madrid Sur. Los resultados de la investigación muestran diferencias y similitudes entre las modalidades bilingües y los centros de enseñanza secundaria. Los resultados conducen a implicaciones pedagógicas y empíricas y se proponen futuras investigaciones.

Palabras clave: Aprendizaje Integrado de Contenido y Lenguas Extranjeras (AICLE), densidad léxica, *academividad*, complejidad gramatical, Funciones del Discurso Cognitivo (FDCs), Lingüística Sistemico-Funcional (LSF).

Table of contents

1. Introduction	1
2. Creation and development of CLIL education	5
2.1. CLIL education in Spain	6
3. Theoretical framework	10
3.1. Language Complexity	10
3.2.1. Lexical density and <i>academicness</i>	10
3.2.2. Grammatical complexity.....	12
3.2. Cognitive Discourse Functions	15
4. Materials and methods	18
4.1. Socio-economic context	18
4.2. Data collection and participants	19
4.2. Methodology	21
4.2.1. Decisions on the analysis and coding of the texts	23
5. Results and discussion	26
5.1. Students' lexical density and <i>academicness</i>	26
5.2. Students' grammatical complexity	30
5.3. Grammar complexity and logico-semantic relations in CDFs	44
6. Conclusion	49
Bibliography	53
Appendix	59

List of Tables

Table 1: Clauses in paratactic and hypotactic clause complexes (Nguyen, 2013: 34).....	13
Table 2: List of Cognitive Discourse Functions (Dalton-Puffer, 2013: 235)	17
Table 3: Data distribution.....	20
Table 4: Comparison of lexical density across bilingual modalities	26
Table 5: Comparison of lexical density across secondary schools	27
Table 6: Comparison of lexical density across MS bilingual modalities	28
Table 7: Comparison of lexical density across MW bilingual modalities	28
Table 8: Comparison of academicness between bilingual modalities	29
Table 9: Comparison of academicness between secondary schools	29

List of Figures

Figure 1: The relations between content, literacy and language (adapted from Cammarata & Cavanagh, 2018, as cited in Morton, 2020:9).....	16
Figure 3: Scheme of taxis (adapted from Halliday and Matthiessen, 2014).....	22
Figure 4: Scheme of clauses classification (adapted from Halliday and Matthiessen, 2014).....	22
Figure 2: Scheme of CDFs (adapted from Dalton-Puffer, 2013).....	23

List of Graphics

Graphic 1: Number of bilingual secondary schools in the Community of Madrid (adapted from Comunidad de Madrid, 2019).....	8
Graphic 4: Parataxis across secondary schools.....	31
Graphic 5: Hypotaxis across secondary schools	31
Graphic 6: Parataxis across bilingual modalities	34
Graphic 7: Hypotaxis across bilingual modalities	34
Graphic 8: Projection and expansion in paratactic clauses in secondary schools	36

Graphic 9: Types of expansion paratactic clauses in secondary schools	37
Graphic 10: Types of expansion paratactic clauses in bilingual modalities	38
Graphic 11: Projection and in hypotactic clauses in secondary schools	39
Graphic 12: Projection and expansion in hypotactic clauses in bilingual modalities	40
Graphic 13: Types of expansion hypotactic clauses in secondary schools	41
Graphic 14: Types of expansion hypotactic clauses across bilingual modalities	43
Graphic 15: Type of taxis elicited by CDFs	45
Graphic 16: Projection and expansion elicited by CDFs.....	46
Graphic 17: Expansion logico-semantic relations elicited by CDFs.....	47

1. Introduction

The field of education has undergone many fluctuations over the years, as it is a discipline that adapts to the changes of society and youth culture (Celce-Murcia, 2021). Nowadays, multilingual and communicative competences have become an indispensable requirement, as living in a globalised world makes intercultural relations crucial. Therefore, languages are the tool necessary to communicate (Putra, 2020).

Communication connects countries, facilitates understanding and accelerates and enhances innovation. The need to establish a common language so that people could communicate with each other, developed into English as lingua franca, i.e., a language for communication used by native and non-native speakers of that language in international contexts (Berns, 2009). It is worth noting that there are other languages spoken all over the world that are used specially for commercial purposes. For instance, in Latin America, Spanish is the language used in cross-border trade, as well as in Middle East trades, the language most spoken is Arabic. However, it was essential to establish a worldwide lingua franca (EF English Proficiency Index, 2021). According to Berns (2009), English as a lingua franca is the most widely used English in the world, as 2.5 billion people of diverse nationalities communicate in English on a regular basis, both in their personal and academic lives. Moreover, a defining characteristic of 21st century society is that people communicate not only face-to-face, but through social media. Thus, English has been expanding and surpassing spheres, as the opportunities to be in contact with this language have been, and are, limitless.

In 2021, the international education company EF ranked 112 countries and regions according to their English skills. The results showed that, in Europe, the level of English is higher than in other regions, with Spain at the Moderate Proficiency level (position 33 out of 112). Although the level of English is improving, this improvement is not fast

enough and needs to be encouraged through education, which is one of the best tools for learning English (EF English Proficiency Index, 2021).

Thus, in recent decades, there has been a great growth in the interest and popularity of bilingual education, also known as CLIL (Content and Language Integrated Content). In the 1990s, both the Council of Europe and the European Commission raised awareness of the benefits of bilingual education with the aim of enhancing the learning of foreign languages, mostly in compulsory education (Coyle, 2008). The term CLIL was coined in this period and set two objectives: to improve the methodology of teaching and learning English as a foreign language; and to foster multilingualism (Llinares & Morton, 2017). According to Marsh (2002:58), CLIL covers any academic activity in which ‘a foreign language is used as a tool in the learning of a non-language subject in which both language and the subject have a joint role’.

Thus, CLIL focuses not only on language teaching, but also on content, so that L2 acquisition and cognitive abilities are integrated and combined. In fact, the first word that form the acronym CLIL is “content”, as academic content guides language learning. It is really complicated to integrate these two aspects, both for teachers and students. For instance, in History subject, learners need to be able to understand events, changes in the past, explain causes and consequences, etc. in a language which is not their native language. Therefore, it is necessary to provide learners with the necessary linguistic resources to express subject content in their L2. Here arises one of the major difficulties of the bilingual approach, as it requires teaching and learning strategies and methodologies different from the traditional ones. Teachers need the ability to adopt a methodology that allows them to introduce and develop activities that combine language and subject content (Pistorio, 2009).

According to Llinares & Nashaat-Sobhy (2021), it is key to analyse and investigate the linguistic difficulties that students face when expressing content in their L2, both in oral and written forms. Thus, this research paper aims to study the linguistic resources

that can help students to establish a link between language and content, so that they would be able to express academic content in their L2.

Regarding students' linguistic expression, one of the main objectives in English as a Foreign Language (EFL) research has been to observe and measure the lexical density of students' texts (Gregori-Signes & Clavel-Arroitia, 2015). Lexical density determines the learner's proficiency in communicating in written form (Dewi, 2017). Not only lexical density will be taken into consideration when studying linguistic proficiency, but also *academicness*, a notion provided by the UAM Corpus Tool (O'Donnell, 2008) based on Averil Coxhead's Academic Word List (AWL). This interest is based on the fact that the study of the *academicness* provides information on students' use of academic terms for effective academic communication. Therefore, in order to study lexical complexity, the lexical density and the *academicness* of the students' texts will be observed in this paper. In addition, grammatical complexity will also be analyzed in this paper, using Systemic Functional Linguistics (SFL).

Halliday and Matthiessen (2014) point out that both cultural and social background are crucial to interpret and understand language. Therefore, SFL is a linguistic model that can be used to connect language and the context in which it is used, language and content in the case of CLIL. According to Llinares, Morton and Whittaker (2012), SFL is directly related to meaning, as people's language choices are determined by social practices such as education. Hence, the logical-semantic relations expressed by students in texts through clause complexes can contribute to the assessment of students' grammatical complexity for the expression of meanings.

Dalton-Puffer (2013) formulated a model that has proved to be useful for helping CLIL teachers and students in terms of language functions used for the expression of academic knowledge. The construct of Cognitive and Discourse Functions (hereinafter CDFs), serves as a bridge between academic language and content knowledge. Dalton-Puffer (2013) states that verbs that describe cognitive operations, such as *describe*, *explain*, or *define* are used to achieve educational objectives. She also points out that these operations

have “specific linguistic realisations, that can be taught to students” (Morton, 2020:8). This research paper will analyse students’ linguistic realizations of Dalton-Puffer’s CDFs.

In this study, I analyse students’ linguistic production taking into account two variables related to students’ context: the socio-economic level of the area in which the secondary school is located, in order to examine whether average per capita income affects students’ performance; and the two types of bilingual education, whether students attend Bilingual Programme or Bilingual Section. The research questions are as follows:

(1a) Does the lexical density and the *academicness* of students differ across bilingual modalities?

(1b) Does the lexical density and the *academicness* of students differ across secondary schools located in areas of different socio-economic level?

(2a) Is the grammatical complexity of the bilingual section students higher than that of the bilingual programme students?

(2b) Is grammatical complexity different across secondary schools located in areas of different socio-economic level?

(3) Are CDFs related to grammatical complexity and the logico-semantic relation between clauses?

Regarding the structure of this research paper, the first section will be dedicated to the presentation of CLIL teaching in Europe and how it has been implemented in Spain. Afterwards, I will present the theoretical framework, in which the SFL approach will be developed, focusing on the importance of lexical density, *academicness* and grammatical complexity. The next part of the theoretical framework will focus on Dalton-Puffer's concept of CDFs. I will then proceed to present the particular context in which this study has been carried out, referring to the different socio-economic levels. Next, I will focus on the methodology and data used for this study. The next point will be devoted to the presentation of the results and the discussion of them based on the concepts studied in the theoretical framework, in addition, the research questions will be answered. Finally, I will

develop the conclusions reached after examining the results. Furthermore, it is proposed that further research should be carried out in order to acquire greater knowledge of language complexity and CDFs in relation to socio-economic level and CLIL education in Spain.

2. Creation and development of CLIL education

For the last three decades, there has been a strong need to boost learners' linguistic proficiency for a number of reasons, globalisation being one of the most important ones. The aim has been to achieve the European Union's objective of mastering two languages. Thus, many teachers, researchers and policy makers have identified CLIL as a key element in achieving this goal (Busse, 2011).

Although the term CLIL was coined in 1994, this approach has a longer history. It is known that the earliest CLIL activities date back to the time when the Akkadians conquered the Sumerians (Mehisto, Marsh & Frigols, 2008, as cited in Adrián, 2011). Sumerian, the local language, was used to teach certain subjects such as botany or zoology to the Akkadians. Another clear example can be found in the Latin language, which for reasons of prestige was the medium of instruction in European universities. More closely related to the CLIL approach are the immersion programmes in Canada, which aimed to teach English-speaking children the official language of Quebec, French. (Adrián, 2011).

According to this author, during the 1980s and 1990s, the term used in both America and Europe was not CLIL, but bilingual education. At that time, states already had a particular interest in improving and promoting the process of language teaching and learning. At this time, the aim was to build Europe as a community, so bilingual education was one of the keys to spreading cultural awareness and the mastering of different languages (Adrián, 2011).

One of the first initiatives that sought to promote innovative methods for language teaching was the European Council Resolution of 31 March 1995. The main objective was to establish the basis for new methodologies on which the education system should

be based in order to break down communication barriers in Europe, thus helping to establish links between European citizens¹.

Although there were European actions that promoted CLIL education such as Erasmus or Socrates and Leonardo Community action programmes, the Council Resolution was the key point for the take-off of CLIL (Adrián, 2011). For instance, in 2003, the Commission launched its Action Plan 2004-2006, in which CLIL was advocated as the main model for achieving the EU's objective of developing language skills. Likewise, the European Language Label Initiative also sought to achieve the goal of "mother tongue+ 2" through CLIL teaching. Another example can be found in the European EuroCLIL network, which brings together researchers, teachers and others seeking to improve language teaching and learning through CLIL (Eurydice, 2006).

The CLIL approach has had great repercussions at European level in recent decades. The same has happened in Spain, where the European objective of language promotion has also been pursued.

2.1. CLIL education in Spain

In Europe, Spain is taking a leading position in CLIL, both in practice and in research, due to its commitment to European policies dedicated to fostering language learning (Coyle, 2010). It is interesting to highlight the situation of the Spanish' education system, as it is a country divided into 17 autonomous communities, plus the autonomous cities of Ceuta and Melilla. Although there is a national law regulating education, the LOMLOE (Organic Law 3/2020, 29th of December), each region enjoys power to administer its education system on the basis of the education law. Thus, across the country, educational models vary significantly from one region to another (Lasagabaster & Ruiz de Zarobe, 2010), with the Community of Madrid being the region of interest in this paper.

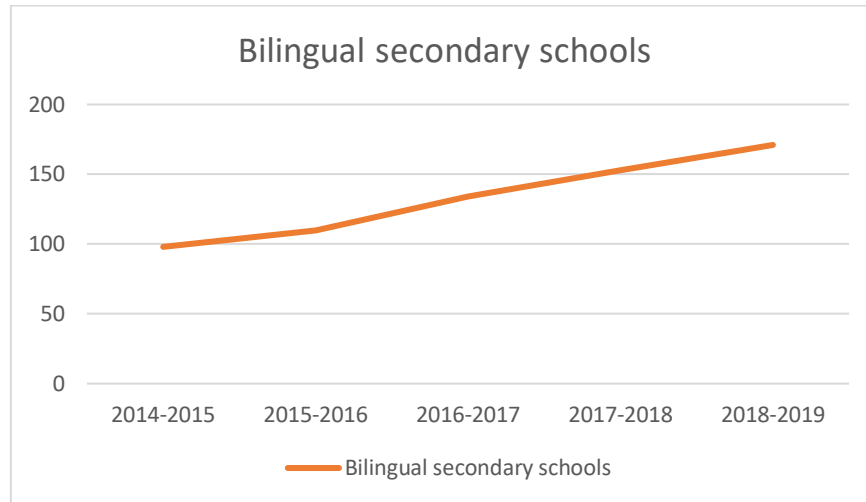
¹ Council Resolution of 31 March 1995 on improving and diversifying language learning and teaching within the education systems of the European Union, Official Journal C207 of 12.08.1995.

Regarding primary education, in 1996, the Ministry of Education, Culture and Sports and the British council signed an agreement to implement a programme known as the MEC/British Council Project, aimed at establishing the Bilingual Project in Madrid and other parts of Spain. The second programme, CAM Bilingual Project, was launched in 2004 by the Comunidad de Madrid (CAM) (Llinares & Dafouz, 2010).

In 2004, the pupils from the MEC/British Council project entered secondary school. In order to prevent students from feeling an abrupt change from primary to secondary education, the project carried out a series of measures. These included showing the pupils the secondary school and introducing them to the teachers, or conversely, inviting secondary school teachers to visit the school to meet their future pupils. Regarding assessment, in the final year of secondary school, the students take the IGSCE (International General Certificate of Secondary Education), to test the proficiency of their knowledge on English language.

For their part, the students that attend the CAM Bilingual Project arrived at secondary education in the 2010-2011 academic year. The measures taken have been related to the training and preparation of teachers and institutions for this new situation (Llinares & Dafouz, 2010). During this year, 32 bilingual schools welcomed students who started the CAM Bilingual Project in 2004-2005 and wanted to continue with this model of education. Now there are almost 200 secondary schools. The following graph shows the evolution of CLIL education in the Community of Madrid from 2014-2015 to 2018-2019. This evolution translates into a significant increase in the number of bilingual secondary schools, since in 4 years the number evolved from 98 secondary schools to 171:

Graphic 1: Number of bilingual secondary schools in the Community of Madrid (adapted from Comunidad de Madrid, 2019)



In Bilingual Secondary, two different tracks are offered: Bilingual Section (hereinafter BS) and Bilingual Programme (hereinafter BP) (Comunidad de Madrid, 2016). BS students have a greater exposure to English. They are taught five hours of advanced English per week. In addition, they receive the tutorial lesson, Social Sciences, Natural Sciences, Geography and History, and Biology and Geology in English. Depending on the availability of teachers at the school, any other subject, except Mathematics, Spanish Language and Literature and Second Foreign Language, can be taught in English (Comunidad de Madrid, 2016). On the other hand, BP students are less exposed to English compared to the students from BS. BP students also receive five hours of English per week and at least one of these subjects has to be taught in English: Physical Education, Plastic and Visual Education, Music, Technology, Ethical Values or other optional subjects. In comparison with Bilingual Section, students from BP are taught Natural and Social Sciences in Spanish (Comunidad de Madrid, 2016).

2.2. Challenges in CLIL education

CLIL education can provide students with many benefits, both linguistic and cognitive. However, there is a lack of clarity in relation to the integration of content and language.

If the balance between these two aspects is not clear, the teaching-learning process will be negatively affected (Morton, 2020).

Thus, one of the main challenges of CLIL in secondary education is teachers' profiles, as they are content experts, but do not necessarily need to be experts in the target language if they have certified knowledge of it. Therefore, one of the main efforts of the administration is focused on the training of teachers in linguistic matters so that they are capable of transmitting knowledge to their students. This challenge is also the focus of attention of CLIL research groups. Currently, there are several research groups from different universities, such as the research group in the University of Alcalá de Henares, the ProCLIL Project in Madrid, coordinated by the University of Cyprus, or the UAM-CLIL Project (Llinares & Dafouz, 2010).

According to Coyle (2008), content teachers must have a high level of English in order to be able to present, discuss and compare information and resources effectively. In order to do so, teachers need to understand the importance of their messages, so they need to organise their discourse and sequence it. Related to this idea is the model developed by Dalton-Puffer (2013), Cognitive Discourse Functions, as it deals with the main academic functions teachers elicit from their students in order to get the most out of them. In addition, Coyle proposes a model of session planning for students to assimilate both content and language. What is proposed is that the teacher plans the lesson on the basis of the 4Cs (Coyle, 2008):

- Content: what is the specific topic of the session?
- Communication: what language should be communicated and how, so that the learners, in turn, know how to communicate the message.
- Cognition: what thinking skills have to be developed and implemented in the session?
- Culture: is there a cultural focus?

3. Theoretical framework

In this section, the theoretical framework underlying the study will be presented. The concepts of lexical density and *academicness* will be studied, as well as grammatical complexity, analysed through the SFL approach. Moreover, the theory of Cognitive Discourse Functions will be introduced, as well as its implications in Bilingual teaching practice.

3.1. Language Complexity

Language complexity is a noteworthy area of research in linguistics. Many scholars have analysed and measured language from various research fields, as in Second Language Acquisition or Computational Linguistics. However, it is difficult to determine the nature of language complexity, as it is a vague concept (Qinghua & Xinxin, 2019).

The aim of this paper is to analyse the language complexity of CLIL students through three linguistic constructs: lexical density, *academicness*, and grammatical complexity when writing about a general topic: *Women Today*. Although this paper has focused on these three linguistic features, another interesting aspect to study in relation to language complexity is lexical variation, i.e., the percentage of different lexical words out of the total number of lexical words.

3.2.1. Lexical density and *academicness*

Lexical density is a widely used measure when working with computer analyses of large data, especially written data, as lexical density is usually higher in such texts than in spoken texts. The term lexical density is the most commonly used to describe the percentage of lexical words out of the total number of words in a text (Johansson, 2008).

According to Johansson, (2008), the major content words are lexical words, which are divided in four categories: nouns, adjectives, adverbs and verbs. Contrarily, those items that do not carry lexical properties are grammatical items, as their function is merely

grammatical-syntactic, they establish the relations between lexical or content words. Therefore, grammatical items are auxiliary verbs, prepositions, pronouns, determiners and conjunctions. Words with lexical values are commonly named as open-ended or open-class words, whereas grammatical items are classified as closed sets. This categorization comes to the fact that, while open class words provide the possibility of including new members of that class into the language, closed set words are unlikely to add new members to their class, as pronouns or prepositions rarely enter a language (Ure & Ellis, 1997, as cited in Johansson, 2008). Although this distinction seems clear, it is not entirely straightforward. Halliday's position towards this classification is reticent, as he claims that an item does not have to consist of only one word. For example, for Halliday (1985), the phrasal verb *turn up* consists of a lexical item, while Ure divides it into a lexical item (*turn*) and a grammatical item (*up*) (Johansson, 2008).

In general terms, the ratio of lexical words in contrast to grammatical items will show the lexical density of texts compared to other texts. Moreover, lexical density, in addition to being easy to operationalise, provides us with “a notion of information packaging; a text with a high proportion of content words contains more information than a text with a high proportion of function words (prepositions, interjections, pronouns, conjunctions and count words)” (Johansson, 2008:65).

In the area of linguistic research, several models of lexical density analysis have been proposed, such as simply calculating the noun density, which consists of dividing the number of nouns in a text by the total number of tokens (Johansson, 2008). However, this paper, thanks to the software UAM Corpus Tool, will look at average number of open class words in each segment, as well at the percentage of words that are open-class. Therefore, a general and specific overview of the lexical density of each group analysed will be obtained.

In addition to lexical density, the percentage of lexical words that are considered academic for the UAM CT will be analysed. In the last few decades, the study of vocabulary

development has become a popular topic of research, as students normally have difficulties when writing in their L2 due to the insufficient knowledge of vocabulary (Nadarajan, 2011). According to this author, many studies have been carried out in relation to this issue, showing that vocabulary proficiency is one of the best indicators of high quality writing skills (Astika, 1993; Santos, 1988 as cited in Nadarajan, 2011).

Academicness is the term used by UAM CT to refer to vocabulary commonly used in academic texts and that reflect users' proficiency of a particular language. This notion is based on Averil Coxhead's Academic Word List. The AWL is composed of 570 word families that occur frequently in academic texts. The list excludes words that appear in the General Service List, that is, the most frequent English Words. It is of great interest to analyse the percentage of lexical words which are on the academic word list, as *academicness* is useful both in oral and written expression, as it is a key point in academic success (Nadarajan, 2011).

3.2.2. Grammatical complexity

The construct of complexity is a dynamic and complex concept that has been widely analysed from different perspectives since the 1990s. According to Housen and Kuiken (2009), complexity, together with fluency and accuracy, figure as a dependent variable which serves as a basic descriptor of performance and proficiency regarding L2. These authors argue that complexity could be defined as the skill of using and combining a wide range of grammatical structures, as well as the ability to use varied vocabulary (Housen and Kuiken, 2009).

In this paper, the SFL framework will be applied, as what is to be studied is students' use of clause complexes, a notion introduced in Halliday and Matthiessen (2014).

A clause is "a group of words, consisting of a subject and a finite form of a verb" (Cambridge Dictionary, 2022). As stated by Eggins (2004), a clause complex is how clauses are associated in sequence, both in oral and written texts. Halliday and Matthiessen (2014) defined clause complex as a logical structure in which clauses are

liked to others, therefore creating relations of interdependency. These authors established that there exist two basic systems that establish how clauses are interrelated: the degree of interdependency and the logico-semantic relation (Halliday and Matthiessen, 2014).

The degree of interdependency is named as ‘taxis’ (Halliday and Matthiessen, 2014). Therefore, clause complexes can be divided into two categories. If the relations between clauses is of independency and equal status, the clauses are paratactic; whereas if there is a dependent and unequal relation between clauses, they are hypotactic (Nguyen, 2013). Thus, the relation between clauses are as follows:

Table 1: Clauses in paratactic and hypotactic clause complexes (Nguyen, 2013: 34)

Taxis	Primary clause	Secondary clause
Parataxis	1 (initiating)	2 (continuing)
Hypotaxis	α (dominant)	β (dependent)

In the parataxis relations (coordination), both clauses are free and independent, as they can function as a whole on their own. One serves to initiate the message, while the following clause continues with this message. The examples of this section are all taken from the corpus analysed in this study:

(1) *“In my grandparent's generation the woman must be in house with kids and now woman will be doing whatever she want”.*

Here we have two clauses. They are both main clauses, as they are equal in status and completely independent. Hence:

||| ¹*“In my grandparent's generation the woman must be in house with kids /and² now woman will be doing whatever she want”.* |||

The contrary occurs in the hypotactic relations (subordination) in which the dominant clause is free, but the independent one is not, as it cannot function as a whole due to its lack of complete meaning (Nguyen, 2013):

(2) *“Feminism is a movement that fight for the rights of women.”*

Here we have two clauses. There is one main clause and another one that is dependent, as it doesn't have meaning on its own. Hence:

||| . ^α*“Feminism is a movement // ^βthat fight for the rights of women”*. |||

When analysing a CLIL text, it is of particular interest to look at the meanings expressed through logico-semantic relations. The logico-semantic system describes and establishes the relation between clauses in terms of meaning. Although there is a wide range of different logico-semantic relations, there are two that are fundamental: expansion and projection. Expansion refers to those clauses that expands the meaning of others, whereas projection means quoting or reporting a clause. These first category is, at the same type, subdivided as follows (Halliday and Matthiessen, 2014):

Expansion:

- **Extension:** one clause develops another by adding something new or giving an alternative. This relation is characterised by the symbol “+”:

(3). ||| ¹*“Women have rights, // ⁺² and they can make a huge impact in the world”*. |||

(4). ||| ^α*“Women were treated as slaves, // ^{+β} whereas men could do whatever they want with them”*. |||

- **Enhancement:** one clause develops another by providing circumstantial features of place, condition, time, manner, cause... This relation is characterised by the symbol “x”:

(5). ||| ¹*“Today women's life change a lot from back then, // ^{x2} they can have a job now”*. |||

(6). ||| ^α*“This movement is celebrated // ^{xβ} because women historically had less rights than men*. |||

▪Elaboration: one clause develops another by commenting, specifying in more detail or exemplifying. This relation is characterised by the symbol “=”:

(7). ||| ¹ “*Usually womens couldn't do anything by her own, // ⁼² get a job, etc.* |||

(8). ||| ^a “*Feminism is an ideology // ^{=b} that aims to achieve women rights*”. |||

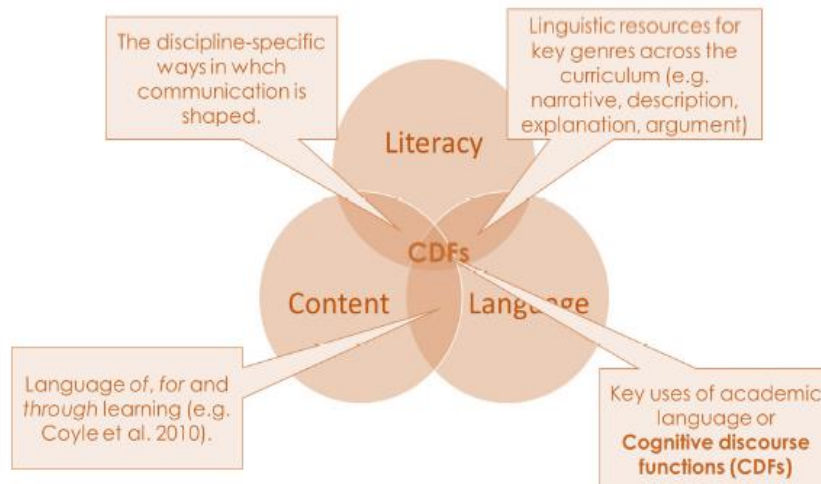
3.2. Cognitive Discourse Functions

As seen in the previous point, one of the main problems in relation to CLIL is the link between content and language. Through the Cognitive Discourse Functions, Dalton-Puffer (2013) establishes a relation between linguistics and education, as the theory analyses aspects of discourse commonly verbalized in the teaching and learning process. Moreover, according to Morton (2020):

“CDFs form a link between cognition and language or thinking and speaking/writing. As such, they are a bridge between content learning objectives, the specific types of communication (literacies) associated with academic subjects, and the language used to express knowledge and thinking.” (p.8)

Llinares & Navarro (2021) support Morton’s statement, as these authors state that even tough the study of CDFs across subjects is a recent area of research, it can lead to solid links between content and language teaching. The following figure extracted from Morton’s work (2020), helps us to visualize the three types of knowledge and how they are related.

Figure 1: The relations between content, literacy and language (adapted from Cammarata & Cavanagh, 2018, as cited in Morton, 2020:9).



The teacher's role does not only involve transmitting content to learners, but it is necessary to specify what learners should do with that knowledge, particularly in relation to cognitive operations such as memorising, analysing or recalling. There is no one single theory that has categorised cognitive operations, but they have been classified in a variety of ways. The best known is Bloom's taxonomy, developed in the 1950s, and later revised by Anderson and Krathwohl in 2001. Although this classification has been useful, one of its major drawbacks is that it can be very messy, as some verbs can describe different thinking skills. The result can be quite disastrous because it can lead to confusion among learners about what to do with the knowledge they are being taught. For this reason, Dalton-Puffer developed the theory of CDFs (2013). Based on the achievement of educational objectives, Dalton-Puffer categorised seven communicative intentions, as seen in Table 2 (Morton, 2020):

Table 2: List of Cognitive Discourse Functions (Dalton-Puffer, 2013: 235)

CLASSIFY	Classify, compare, contrast, match, structure, categorize, subsume
DEFINE	Define, identify, characterize
DESCRIBE	Describe, label, identify, name, specify
EVALUATE	Evaluate, judge, argue, justify, take a stance, critique, recommend, comment, reflect, appreciate
EXPLAIN	Explain, reason, express clause/effect, deduce
EXPLORE	Explore, hypothesize, speculate, guess, estimate, predict
REPORT	Report, inform, recount, narrate, present, summarize, relate

Generally, CDFs are not consciously organised and planned by teachers. However, in CLIL education there is a demand related to communicative needs, as it is important for students to understand thought processes and verbalize them (Llinares & Navarro, 2021) By having the academic language organised according to Dalton-Puffer’s CDFs, CLIL curriculum developers, teachers and students would have access to the same terminology. Therefore, confusion and ambiguity would be eliminated as the communicative intentions would be clear and “language-based interventions would be much more focused on the content-learning and literacy needs” (Morton, 2020: 11).

According to Llinares & Navarro (2021), CDFs can guide students thinking towards their writings and speech acts. Moreover, CDFs are observable and can shed light on the relations between CLIL and the teaching and learning process of content-subjects. Numerous studies have therefore been carried out in relation to CDFs (Whittaker & McCabe, 2020 on *evaluate*; Llinares & Nashaat-Sobhy, 2021 on *define* and Evnitskaya & Dalton-Puffer, 2020 on *classify/compare*).

4. Materials and methods

In this section, the context of this research will be introduced. In addition, the data used, as well as the software employed to analyse these data will be presented.

4.1. Socio-economic context

In the present study, five groups of students of 4th year of Compulsory Secondary Education were given a prompt to write a text where lexical density, *academicness* and grammatical complexity were measured. These groups belonged to three public bilingual secondary schools, located in three localities in the Community of Madrid, Spain. Three groups were scholarised in two bilingual secondary schools in Madrid West, whereas the other two groups belonged to a bilingual secondary school in Madrid South.

It is noteworthy to highlight the socio-economic level of each secondary school, as it is a variable that will be considered when analysing the results. Moreover, these data are also necessary to answer research questions 1 (b) and 2 (b), since one of the hypotheses of this study is that socio-economic status directly affects students' performance.

According to the Instituto Nacional de Estadística (2020), the household net income of residents in Spain in 2020 was €30,690. However, the data provided by the same source show that the average income per person in the Community of Madrid is higher than the national average, exceeding it by €6,361 (INE, 2020).

Two of the three secondary schools that are analysed in this study share socio-economic characteristics and will therefore fall into the same group. We refer to these two schools as Madrid West (MW) secondary schools. According to the data provided by the *Atlas de distribución de renta de los hogares* (INE, 2019), the average income per household of Madrid West secondary schools is 51,986.86€ and 60,945.58€. Their most important economic sector is Business and financial services (Comunidad de Madrid, 2022). Regarding the *Report on registered unemployment per locality* by the Regional Ministry of Economy, Employment and Competitiveness (2019), these two areas had

2,917 and 1,436 unemployed people. According to the INE (2019), the unemployment rate in the Community of Madrid is 9,99%. The unemployment rates of the two localities in which the Madrid West Secondary schools are located are below the average (5.59% and 4,90%) (Expansion-Datosmacro, 2019).

The third secondary school, referred here as Madrid South (MS) secondary school, belongs to a locality in a lower socio-economic area. Although the average income per household is higher than the average for the Community of Madrid, the difference with the Madrid West secondary schools is still noticeable. According to INE (2019), the average income per household in this locality is €42,790.63. Regarding the economic sectors, the most important one in the locality is the Distribution and hospitality services (Comunidad de Madrid, 2022). The unemployment rate in this locality is 10.98% (Expansion-Datosmacro, 2019), as 10.899 inhabitants are unemployed; thus, the Madrid South secondary school is located above the average unemployment rate in the Community of Madrid (Regional Ministry of Economy, Employment and Competitiveness, 2019).

4.2. Data collection and participants

The data used for this paper belong to the UAM-CLIL research group². The texts from the Madrid West secondary schools were collected during the years 2018-2019 and 2020-2021, and the data from the Madrid South secondary school date from the academic year 2021-22.

These data contain texts from students from 4th year of Compulsory Secondary Education (15-16 years old). The total number of texts is 88, 22 texts belong to each group. The data from the MWBP group consisted of only 22 texts, while there were more texts collected from the other groups. The selection of the latter texts has been done at random in order to analyse the same amount of data in all groups. Therefore, the total

² LongAd-CLIL

number of texts (88) is distributed among 5 groups divided into 3 secondary schools. Of the 5 groups that are analysed in this paper, 3 belong to the Madrid West secondary schools. That is, within the western secondary schools, one BP group and two BS groups are analysed. This is due to the lack of material of one of the BS groups. Therefore, this problem has been solved by combining data from two secondary schools with very similar characteristics. Hence, the Madrid South secondary school provides two groups to be studied, one from BS and the other from BP. These data are summarised in the following table:

Table 3: Data distribution

Madrid South secondary school		Madrid West secondary school 1		Madrid West secondary school 2	
Bilingual Section	Bilingual Programme	Bilingual Section	Bilingual Programme	Bilingual Section	Bilingual Programme
22 texts	22 texts	18 texts	0 texts	4 texts	22 texts

The data analysed were collected during the English class. As previously mentioned, the difference between BS and BP students lies in the number of hours that they are in contact with English per day. These students were given a prompt on feminism (see appendix) in which the 7 CDFs are present, in order to get them to write a text on a topic they had not dealt with in class but on which they should have ideas and a clear opinion.

The aim of this prompt is that the students are able to write a structured text by answering all the questions. The words that appear in bold in the prompt refer to Dalton-Puffer's (2013) construct of CDFs. Thus, students are asked to *define, explain, describe, compare, evaluate, explore* and *report* on various aspects related to feminism. It is possible that the CDFs imply a cognitive difficulty, but it is believed that at this age the cognitive development of the students is sufficient to answer this prompt in their L2.

Furthermore, Vygotsky (1978) states that every learner moves between two dimensions: the one in which the learner is able to fend for himself, and the one in which the learner is unable to do so. Between these two zones is what Vygotsky called the "zone of proximal development", a dimension in which the learner is able to complete tasks if he is helped and guided to do so. Thus, the objective is focused on this zone of proximal development, where the learner will achieve a higher level of development. In addition, students will not be engaged if they do not feel challenged.

4.2. Methodology

In the last decades, linguists have become more and more interested in linguistic patterns that needed the help of specific software. Thus, different software has been developed to help with the annotation of text corpora. Therefore, in order to accomplish the aim of examining students' responses to CDFs, grammatical complexity and lexical density and *academicness*, the UAM Corpus Tool (hereinafter UAM CT) has been used (O'Donnell, 2008).

UAM CT allows to create user-defined layers that enable to define a hierarchy for this specific layer. In addition, it is possible to divide the text into segments, encoding each segment according to the layers that suit it, both at semantic-pragmatic and syntactic levels. This tool also provides automatic tagging based on lexical patterns (O'Donnell, 2008). Moreover, lexical density, textual complexity or subjectivity can be analysed statistically and compared through this software.

Firstly, the texts have been classified in two main groups, which in turn have been classified in two subgroups. Therefore, half of the texts belong to Madrid South, whereas the other half belongs to Madrid West. Both groups are subdivided into bilingual programme and a bilingual section. Thus, the scheme is as follows:

- MS-BP-Tⁿ
- MS-BS-Tⁿ
- MW-BP-Tⁿ
- MW-BS-Tⁿ

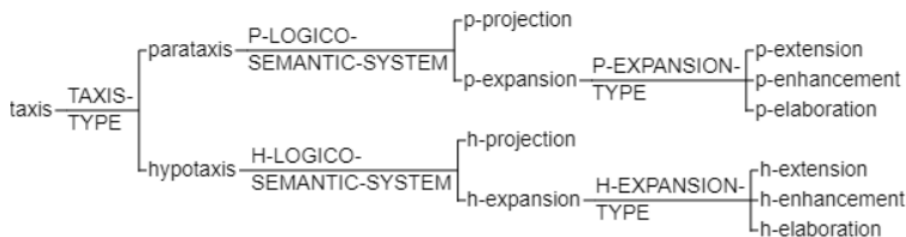
The analysis of grammatical complexity has been made according to Halliday's scheme of taxis and logico-semantic relations. In order to do so, the texts have been coded manually. As Halliday and Matthiessen (2014) established, clauses can be divided into two categories depending on the degree of interdependency. Therefore, the first manual annotation of texts segments is based on the dependency of their clauses (parataxis and hypotaxis).

Figure 2: Scheme of taxis (adapted from Halliday and Matthiessen, 2014)



In addition, the relation between clauses in terms of meaning are studied. Thus, once the differentiation between parataxis and hypotaxis has been made, clauses are divided into *expansion* and *projection*, depending on whether clauses extend the meaning of others, or whether they quote or report a clause. Hence, if the sentence shows an expansion of meaning, it will be studied whether the relation is of *extension*, *enhancement* or *elaboration*. Therefore, the scheme developed with UAM CT is as follows:

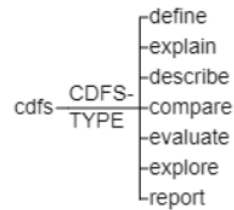
Figure 3: Scheme of clauses classification (adapted from Halliday and Matthiessen, 2014)



Moreover, one of the objectives of this paper is to analyse how students respond to CDFs in terms of grammatical complexity and the logical-semantic relation between

clauses. For this purpose, a layer has been created in UAM CT named CDFs, making a distinction between each of the 7 types. Then, each students' text has been manually annotated, indicating the segments that corresponded to each CDF elicited by the prompt.

Figure 4: Scheme of CDFs (adapted from Dalton-Puffer, 2013)



Once the texts have been coded regarding CDFs, grammatical complexity and logico-semantic relations, the results will be presented in graphs created by UAM CT. In addition, a more qualitative analysis of the results will be carried out in order to justify and illustrate certain linguistic phenomena.

4.2.1. Decisions on the analysis and coding of the texts

As mentioned previously in this paper, the texts that are part of the corpus of analysis have been chosen at random, so as not to create any bias. However, when it came to classifying and annotating, there were two texts that posed such difficulties that they were eliminated from the corpus and replaced by others, also chosen at random. The impossibility of analysis was due to the lack of coherence and cohesion of the clauses, which meant that they were not valid for the study. It is worth mentioning that both texts belonged to the Bilingual Programme group of the Madrid South secondary school.

Throughout the study, we found similar situations but not as critical and conclusive as the previous one. Occasionally, perhaps due to a lack of proficiency in L2 writing skills, the students alternated sentences which could be classified according to Halliday and Matthiessen's (2014) model of clause complexes, with others which were impossible to annotate due to the lack of meaning. In these cases, the decision was not to eliminate the

text from the corpus, but to classify and code those sentences which did present coherence and cohesion, leaving aside those which did not allow any kind of analysis. An example of this can be found in the following text:

(9) *“Feminism is a lot of girls fight the woman have the same right with the man.*

I think the mand and the woman same pained.

*Yes **because** the girls is a person **and** have the same rights.*

*The womans prefer life today **because** the other generation the womans not have rights.”*

In example (9), the first two clauses were not coded (*“feminism is [...] man”/ “I think [...] pained”*). However, the last two, although they present grammatical errors, are linked by logico-semantic relations of *enhancement* and *expansion* and have therefore been coded and analysed in the present study.

Furthermore, in the following example, we can see that the clause is incomplete as the first part is elliptical. Based on the prompt given to the students, we understand that the first part of the clause would be: “today, there is a women’s movement because (...)”. Therefore, we have classified this relation between clauses as *hypotactic*, *expansion*, *enhancement*, as the student is expressing a relation of *cause*.

- (10) *“**Because** is necessary for the right of the women’s.”*

The students' punctuation also posed a challenge in classifying the logico-semantic relation of their clauses, as they sometimes did not place punctuation marks correctly. There have been two different scenarios related to this issue. In example (11) the clause starts with the connector “*and*”, but it is a clause completely independent. It is coherent and cohesive, but it functions on their own. In fact, the student has separated the clause from the previous one by writing it in a new paragraph. Thus, this sentence has not been classified as having a logico-semantic relation, as it functions independently. However, the contrary happens in example (12). In this case, the student has placed a full stop in the middle of the sentence, therefore splitting up two sentences that are actually linked by

a parataxis relation of extension. This case is different from example (11) for two main reasons. The first one being the relation between clauses, as, although both sentences can act independently, there is a clear relation of parataxis in example (12); and, secondly, because the student has written the connector "but" in lower case and the clause continues in the same paragraph, which leads us to interpret this as a simple punctuation error.

- (11) *"my grandmother enjoyed a pretty equal life, but for most people it wasn't that way. **And** that leads me to the next point, the owmen's movement we have in Spain is not being done properly."*
- (12) *"In 8th of march the women's revelation in the streets. **but** something women are crazy."*

On some occasions, the students' texts were grammatically incorrect, but they established logical-semantic relations and maintained coherence and cohesion across their clauses (example 13). As Llinares & Navarro (2021) did in their study, in the cases where students didn't show grammatical accuracy, I have prioritised the logico-semantic relation between clauses and dismissed grammatical rules. Another case related to grammatical errors is found in example (14), which show the confusion between the prepositions "to" and "for" to express purpose.

- (13) *"Men aren't do nothing **or** weren't fight for this day or for women."*
- (14) *"**For** stop the violence, the women's need to equal representation in the jobs, compare with others."*

In addition, following the scheme mentioned above, it is interesting to mention that students often confused homophone words such as "were" and "where". As in the previous examples, the same procedure has been followed. Therefore, in order to carry out the analysis, even if there were errors in students' production, the logico-semantic relation between clauses have been classified as *hypotaxis-expansion-enhancement*:

- (15) *"The 8 of march people do a manifestation **were** a lot of people go to revindicated the womans (derechos)"*.

5. Results and discussion

The aim of this research paper is to assess students' language proficiency. Furthermore, it is of great interest in this study to observe the different language performance of students attending the two modalities of bilingual education (BP and BS), as well as to examine the proficiency of students attending secondary schools located in two areas of different socio-economic status. In order to achieve this aim, three different analyses have been carried out. Thus, in this section we will look at lexical density and *academicness*. The second analysis is related to grammatical complexity, i.e., the type of taxis the students use in their clauses and the relations they establish between them. In addition, we will examine which taxis each Cognitive Discourse Function generates. In other words, the third aim is to study whether students use different grammatical complexity or logico- semantic relations according to the CDF they have been given.

5.1. Students' lexical density and *academicness*

The prompt proposed to the students deals with a general topic which had not been specifically taught in English class or in any other subject. Thus, we are interested in analysing both lexical density and *academicness* in order to see whether BS students apply their advanced knowledge of the language to non-specific texts which do not appear in the English curriculum or in the curriculum of any other subject that they are taught in English. In addition, the aim is also to study whether the socio-economic level of the area where the school is located affects the students' language proficiency.

The table below establishes the comparison between the lexical density of the students from Bilingual Programme and from Bilingual Section:

Table 4: Comparison of lexical density across bilingual modalities

Lexical Density	bilingual-programme	bilingual-section
Lexemes per segment	52.66	87.5
Lexemes % of text	44.49	43.58

The results of this analysis are surprising and very interesting as they do not show a significant difference between the two groups. The students belonging to bilingual section show a higher use of open-class words per segment, 87.5 versus 52.66. In comparison to the bilingual section group, the percentage of lexical density of students from the bilingual programme is higher. Although it would be expected that the lexical density of students from bilingual section would be higher in the two categories provided by UAM CT, the difference is not significant enough to draw conclusions. In fact, the p-value that has been calculated by UAMCT is 0.2939, that is, higher than 0.05, which means that the null hypothesis, that bilingual section texts will be more lexically complex has not been fulfilled, as the results are not sufficiently significant.

To continue with the analysis, the lexical density of the Madrid West secondary school and Madrid South secondary school groups has been compared. The interest lies in the question of whether the lexical difference between the groups in this study is significant. Thus, the following table shows the data extracted by UAM CT:

Table 5: Comparison of lexical density across secondary schools

Lexical Density	madrid-south	madrid-west
Lexemes per segment	72.93	68.16
Lexemes % of text	43.2	44.77

In this case, although the total number of lexical words in the Madrid South group is higher, 72.93 versus 68.16, the percentage of lexical words is lower, by 1.57 points. The p-value extracted by UAM CT is 0.0620. In this particular case, the null hypothesis is that students in the Madrid west group should show a higher lexical density than those students in the Madrid south secondary school. Although this hypothesis is satisfied by the results, the p-value indicates that the results obtained are not sufficiently significant to verify the hypothesis.

In order to find conclusive results, an extra comparative analysis was carried out. In this case, the lexical density of students in the bilingual section and bilingual program in MW and the same groups MS was compared. However, as shown in the following two tables, the results are not significant either, in both cases the p-value is higher than 0.05, so no conclusions can be drawn regarding lexical density.

Table 6: Comparison of lexical density across MS bilingual modalities

Lexical Density	MSBP	MSBS
Lexemes per segment	44	101.86
Lexemes % of text	44.86	42.52

Table 7: Comparison of lexical density across MW bilingual modalities

Lexical Density	MSBP	MSBS
Lexemes per segment	62.95	73.14
Lexemes % of text	44.32	45.15

As mentioned in the theoretical framework and according to Johansson (2008), the study of lexical density is equivalent to the analysis of the number of nouns, adjectives, adverbs and verbs out of the total number of words in a text. The explanation drawn from the results of this analysis is that the analysis of lexical density is not accurate for this research as learners may have systematically repeated the same words belonging to these four grammatical categories, or used simple words which do not require specialised or advanced knowledge of the English language. It is for this reason that *academicness* has also been taken into account to analyse the students' lexical proficiency.

Therefore, the same study has been carried out but focusing on *academicness*. The table below shows the *academicness* of the students from Bilingual compared to the *academicness* of bilingual section students.

Table 8: Comparison of academicness between bilingual modalities

Academicness	bilingual-programme	bilingual-section
Academic Word Use	3.73%	4.83%

In this case, the results are significant, with a p-value of 0.0022. Therefore, it is possible to appreciate certain difference in terms of *academicness* between BP students and BS students, who are in contact with the English language during a longer period of the school timetable. Thus, at least in this case in which bilingual modalities are compared, we can deduce that although the p-value related to lexical density is not significant, it may be due to the fact that BP students make use of words which are neither academic nor specialised. The texts from BS students, although they are mostly not composed of academic words, show a greater command of the specialised language, presumably thanks to the hours they spend in contact with the L2 in academic subjects such as history, biology, etc.

In fact, it is worth noting that students from BS, especially those from Madrid-West secondary school, use academic connectors to start developing their ideas or to link clauses. An example of the above-mentioned academic connectors can be seen in the following extract from a text produced by a MW-BS student:

(16) “*Taking this into account, I think that [...]*
Even though they can do it, it’s a fact that [...]
At last, it’s important for you to know [...]”

The following table shows the *academicness* of students from the Madrid South secondary school compared to the students from Madrid West secondary school:

Table 9: Comparison of academicness between secondary schools

Academicness	madrid-south	madrid-west
Academic Word Use	4.39%	4.47%

Contrary to the previous case, the comparative analysis between schools of different socio-economic status does not show significant results, as the p-value is 0.8185. Students from the school with the highest socio-economic status show a higher percentage, but there is not a significant difference.

To conclude, and with the aim to answer to the first research question, it can be deduced from the analysis that the study of lexical density in this paper did not yield significant differences across groups, probably because this analysis takes into account all lexical words (nouns, adverbs, adjectives or verbs), even if they are repeated throughout the text or are basic words of the English language. Therefore, it cannot be said that we have found relevant differences between the four groups of study.

On the other hand, after the analysis of the students' use of academic expression, we can conclude that there is no significant difference between secondary schools in areas of different socio-economic status. However, BS students, in comparison to BP students, have shown that they are able to use their advanced knowledge of their L2 to write about a non-specialised topic such as feminism. Thus, it can be stated that part of the hypothesis is confirmed, as we have found significant differences between two groups of study, BS and BP. However, the socio-economic level has no relevance in terms of lexical density and *academicness*. Hence, the most significant variable regarding these results is the number of hours that the student is in contact with English during school hours and/or the type of exposure (academic English in science, history, etc.).

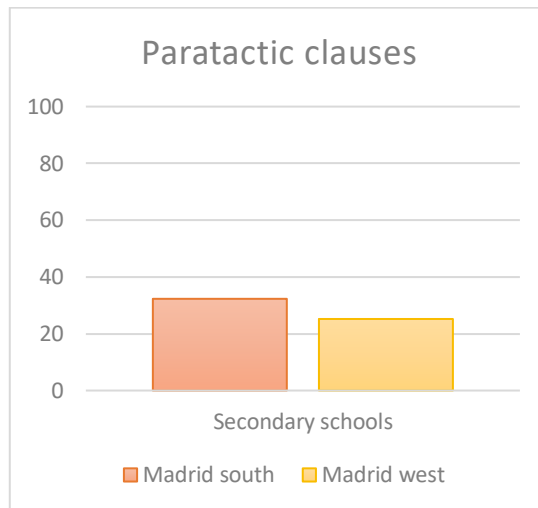
5.2. Students' grammatical complexity

In this section we present the results on the different types of clauses complexes that students from 4th grade of Compulsory Secondary Education used in their writing, following Halliday and Matthiessen (2014).

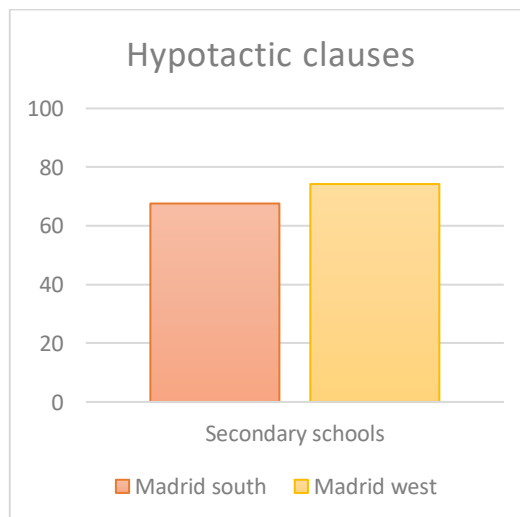
The following graphs compare the degree of interdependency of students' clauses comparing schools from different socioeconomic areas. Therefore, the contrast of the use

of parataxis and hypotaxis by the students from the Madrid South secondary school compared to that by the students from the Madrid West secondary school is shown below:

Graphic 2: Parataxis across secondary schools



Graphic 3: Hypotaxis across secondary schools



As it can be seen above, students from the Madrid West secondary schools make less use of parataxis (25.2%) than those from Madrid South secondary school (32.3%). The

p-value calculated by UAM CT is 0.0112, therefore the results are significant. Likewise, the results extracted from the use of hypotaxis to join clauses are also significant and conclusive. Students from the Madrid west secondary schools make greater use of this type of relation between clauses, 74.3% of their clauses are linked by a hypotactic relation, while in the case of the students from Madrid south secondary school, this percentage is 67.6%. Moreover, the p-value for this result is also 0.0112, so we conclude that the results are significant.

According to Eggins (2004), texts have to be structured and their clauses have to be linked to each other so that recipients can interpret the speaker's message and intention. She states that there are conjunctions and connectors which indicate what kind of taxis has been used to link clauses. Thus, in the case of parataxis, the connectors that indicate this type of relation are *and, but, or, neither... nor*. The most common conjunctions which establish hypotactic relations are *because, while, when, if...* as well as relative pronouns (*that, which, who*), which serve to specify, comment on or exemplify the previous clause.

In the following extracts we find some examples of paratactic clauses produced by the students from Madrid South secondary school:

- (17) “Women have rights **and** they can make a huge impact in the world **but**, in my opinion we need men's help [...]”
- (18) “Of course, also men are killed **and** raped, **but** not for the same reason as women are.”

Likewise, in the following examples we can appreciate the use of the hypotactic relation to link clauses. These are extracts taken from the Madrid West secondary school students:

- (19) “In the past, women used to work at home **while** they were taking care of her children.”
- (20) “**When** my grandmother was young she was the one **who** cooked, cleaned and **if** my grandfather had to move houses because of it's job she was the one **who** had to adapt to my grandfather's conditions [...]”

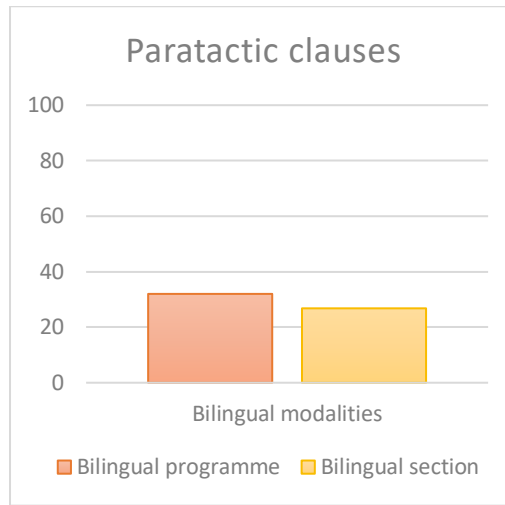
The use of hypotaxis requires a greater knowledge of the language, since through hypotactic connectors, relations of causality, consequence or condition are established; in addition, through these connectors, the meaning is expanded by adding an idea related to the first clause. On the contrary, the use of parataxis is simpler and doesn't involve a high level of proficiency in English, as paratactic connectors do not allow a relation that goes beyond the linking of clauses to add to or contradict an idea. Therefore, the conclusion that can be drawn is that in this case, the socio-economic level could play a role in the performance of the students, since results have shown that there are significant differences that indicate that students from schools with a higher socio-economic level have a greater command of the hypotactic connectors.

Although it is not the object of study in this research paper, I think that it is interesting to note that on numerous occasions, especially in the case of the texts produced by the students from Madrid south, the students made use of juxtaposition. That is to say, they do not use either parataxis or hypotaxis connectors to join their clauses but link them by means of commas. There are some possible explanations for this recurrent phenomenon, as students may not have the level required to use academic connectors to link their sentences or they may not have sufficient knowledge of the language; another possible explanation is that students have not had sufficient training in writing. The following example has been extracted from MSBP:

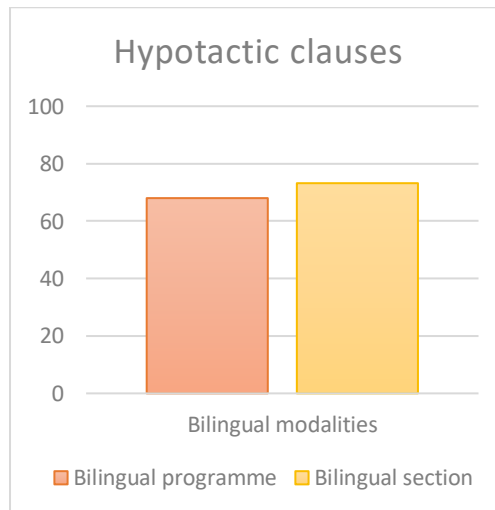
(21) "Yes, i think this movement in Spain benefittinf the society, we need fight for the right, for stop the violence, the women's need to equal representation in the jobs, compare with others."

The following graphs also analyse the use of hypotaxis and parataxis but in this case the comparison is between BP and BS students:

Graphic 4: Parataxis across bilingual modalities



Graphic 5: Hypotaxis across bilingual modalities



Whilst the p-value extracted by UAM CT shows that the results are not significant (P value 0.0733), some conclusions can be drawn from the percentages shown in the graphs. As it can be observed, it is noteworthy to state that the use of hypotaxis is higher in the case of BS students (73.2% in comparison to 68.0%). Contrarily, the number of paratactic

clauses in the case of BP students is higher than in the case of BS students (32.0% opposed to 26.8%). However, although the percentage of parataxis is lower in BS students, the total number of paratactic clauses is higher (189 compared to 121 in the case of BP students). This may be due to several factors. Firstly, it is possible that the texts produced by BS students are longer, which makes it necessary to use a greater number of connectors to link the clauses. Another possible explanation is related to what was mentioned above concerning juxtaposition. Even if the texts were approximately the same length, the BP students may have employed the use of commas to link their clauses instead of connectors indicating parataxis or hypotaxis, thus showing a smaller number of both paratactic and hypotactic clauses.

Example (22) illustrate the use of paratactic and hypotactic clauses by students from BP:

- (22) *“The movement in Spain might be not as big as somewhere else, **but** it is very important **because** it helps us to inform people about feminism.”*

Correspondingly, let us now examine how BS students link their clauses:

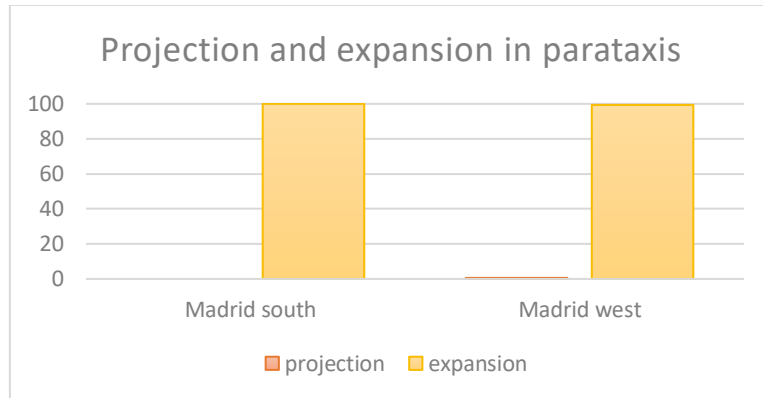
- (23) *“I really think **that** this movement is benefiting society **because** each day **that** passes more rights are being achieved **and** the ideology of people is changing.”*

Although both groups make use of both paratactic and hypotactic connectors, the BS student shows a greater command in his/her English expression, since in example (23), the student has written 4 connectors, 3 of them hypotactic (2 of elaboration “*that*” and 1 of enhancement “*because*”). The BP student, although he/she has linked the clauses with “*but*” and “*because*”, has not elaborated on his clauses by adding extra information.

The following graphs focus on the logico-semantic relations between clauses. It is of great interest in this paper to analyse which type of relations are the most used by, on the one hand, BS and BP students and, on the other hand, by students from Madrid west and Madrid south secondary schools.

The analysis of paratactic sentences is divided into two graphs. Firstly, the following graph shows the percentage of clauses related through *projection*, or *expansion* across socioeconomic areas.

Graphic 6: Projection and expansion in paratactic clauses in secondary schools



As it can be observed, the students' use of projection is much lower, even negligible than that of expansion (0% in the case of Madrid south students and 0.4% in the Madrid west students' clauses). In this case, there is no difference between students from different secondary schools. What is interesting to note is the difference between projection and expansion. This significant difference might be due to the channel in which the students have been asked to respond to the prompt. In spoken texts, projection may have been more used to express directly other people's thoughts. However, in written texts it is difficult to introduce the exact words of other people as the prompt is asking for the students' thoughts, ideas and impressions. The only clause in which a student has introduced direct speech has been in the MS-BS group:

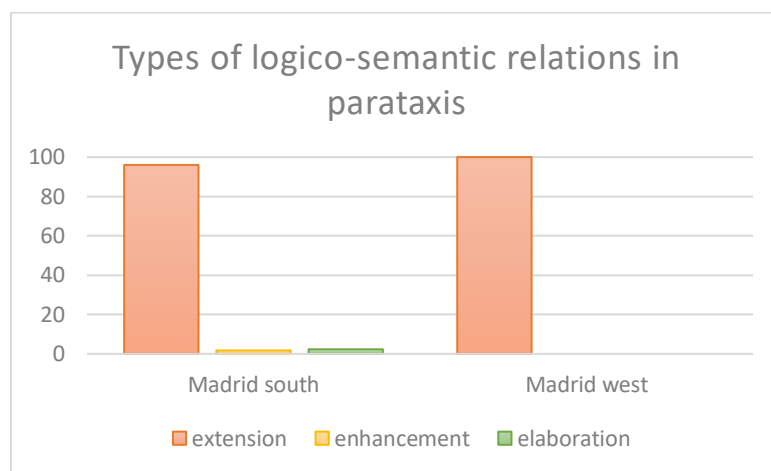
- (24) *'The history will be war after war as long as women hasn't say "stop, we are here and we are people with opinions".'*

It is of no interest to include the graphic of the same data but comparing the two bilingual modalities since the only case in which a student has written paratactic clauses linked by projection is in the example mentioned above. The rest of the students have

chosen not to use the direct speech to report their thoughts, probably due to the characteristics of the prompt.

The next section focuses on the paratactic relations between clauses, more specifically on the type of relation within paratactical expansion. That is, the use by BP and BS students, as well as by MW and MS students, of the logico-semantic relations of paratactic extension, enhancement and elaboration will be compared.

Graphic 7: Types of expansion paratactic clauses in secondary schools



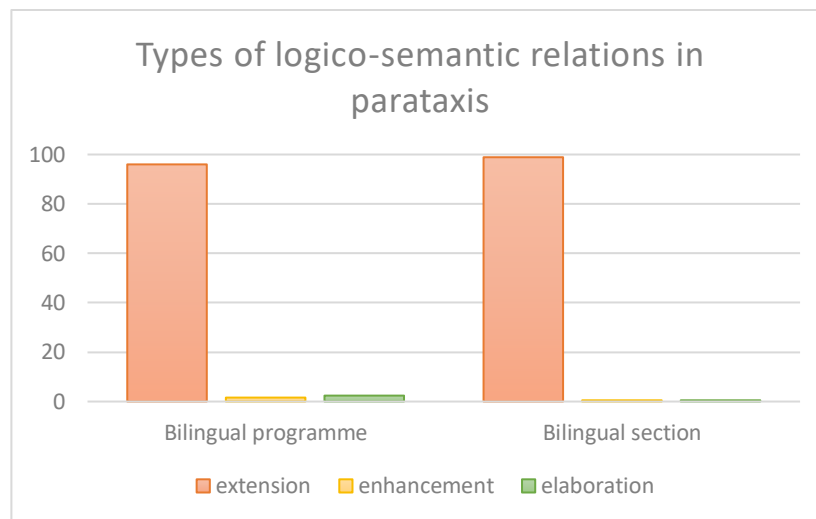
As seen above, the use of parataxis may indicate less language proficiency as it involves linking sentences with simpler and less academic connectors. Moreover, the logico-semantic relations of enhancement and elaboration in parataxis are less common than those of extension, since they do not require connectors, but are independent sentences linked by meaning. While it is true that the use of paratactic enhancement (1.7%) and paratactic elaboration (2.3%) occur in isolated cases, it is interesting to see that these cases occur only in the lower socio-economic secondary school. One possible explanation is that these students did not possess sufficient linguistic skills to relate sentences by means of connectors. On the other hand, the students at the Madrid West secondary school do not make use of this type of logico-semantic relations but use connectors such as *and* or *but*. In the following sentences we find some of the examples

in which students have used the logico-semantic relations of enhancement and elaboration. These examples belong to students from MSBP:

- (25) *‘Today women's life change a lot from back then, **they can have a job now, vote, decide** by her own [...]’.*
- (26) *“[...] usually womens couldn't do anything by her own, **get a job, etc.**”*

The graph below compares logico-semantic relations in parataxis now between students from bilingual programme and bilingual section:

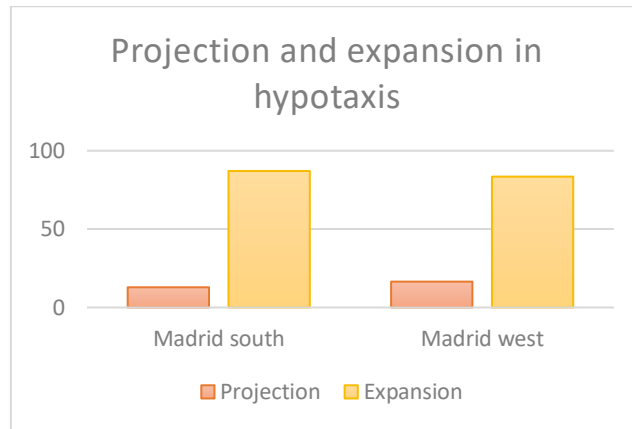
Graphic 8: Types of expansion paratactic clauses in bilingual modalities



The seven cases in which students have established logical-semantic relations of enhancement or elaboration are divided between BP and BS students. However, it is worth mentioning that this division is not exact, but in the case of the BP texts they account for 1.6% and 2.4% while in the case of the BS students the percentage is only 0.5% in both enhancement and elaboration cases.

Once the paratactic logico-semantic relations across the four groups have been studied and compared, let us see graph 11 to analyse the percentage of clauses linked by projection and expansion across bilingual secondary schools:

Graphic 9: Projection and in hypotactic clauses in secondary schools



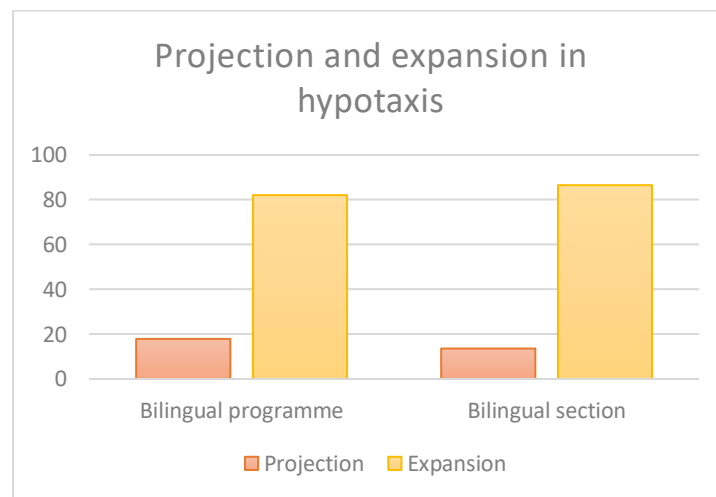
When students use the logical-semantic relation of projection to link their clauses, they usually intend to introduce an idea or thought through reported speech. In the case of the comparison between Madrid south and Madrid west secondary schools, the percentage of projection is 12.9% and 16.5% respectively. In absolute value, the percentages represent, in the case of Madrid west, 68 of the total number of semantic logical relations (412), and in the case of Madrid south, 46 out of 356. Expansion is the most commonly type of relation used by students for several reasons. Mainly, expansion covers more possibilities, since it encompasses relations of extension, enhancement and elaboration. Moreover, the prompt given led students to establish circumstantial complement relations, especially of causality (because). In the cases where students have implemented projection, the introductory verb most frequently used was *to think*, as it is a simple verb through which they express their own ideas as well as those of society. The study carried out by Llinares & Navarro (2021) on clause complexes and CDFs is worth mentioning as their results also showed that the use of “*I think*” was remarkable in the case of hypotactic projection of ideas. Despite the large number of synonyms they could use to make their text more academic (imagine, expect, believe), students generally prefer to stick to the verb *think*. Some examples from MSBP to illustrate this phenomenon are:

It is also remarkable how students expressed content through the hypotaxis projection of ideas. In other words, they used the expression I think that... to put into words their thoughts about a particular imaginary situation.

- (27) “*I **think** is same the life before, the generation in my grandparents, compare with now, but the life now is a little worst [...]*”.
- (28) “*But today I **think** this movement is not necessary [...]*”.

The following graph shows the comparison of projection and expansion within hypotactic clauses across the two bilingual modalities:

Graphic 10: Projection and expansion in hypotactic clauses in bilingual modalities



As for BS students, the use of projection is 13.6% (70 out of the total of 516 relations established in their texts). The percentage of projection in the case of BP students is 17.9% (46 out of a total of 257 relations established in their texts). Although there is no significant difference between these percentages, it is noteworthy to highlight the contrast in the number of logico-semantic relations established between the two groups, 516 as opposed to 257. This may be due to the fact that BS students are more used to writing academic texts in English, thus their productions are considerably longer, so that as they write more, they have to establish more relations between their clauses. Moreover, as has

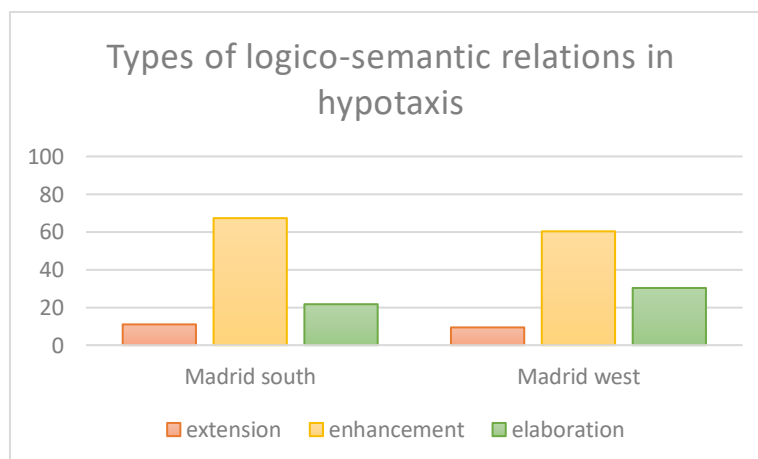
been mentioned throughout this paper, BP students tend to establish more paratactic and juxtaposition relations between their clauses, so the number of hypotaxis is lower.

Another interesting point is related to the introductory verbs of reported speech. In the previous comparison between secondary schools of different socio-economic levels, no significant differences have been established since, although the most used introductory verb was *to think*, on exceptional occasions students from both secondary schools used synonyms to report their ideas. However, it is worth noting that all the examples of other introductory verbs belong to the bilingual section. Thus, in order to give their texts greater complexity and lexical diversity, BS students introduce clauses with verbs such as *believe*, *suppose* or *hope*. Some specific cases are shown below:

- (29) “*I **hope** that in sport for example the level of the men is higher than the level of the women.*”
- (30) “*It was **supposed** that their only objective was to serve their husband.*”
- (31) “*I **believe** that this is not feminism at all.*”

Having introduced the graphs concerning the percentage of use of projection and expansion, the following graph shows the data related to the logical-semantic relations of extension, enhancement and elaboration within the hypotactic relations. Firstly, the data extracted from the Madrid west and Madrid South secondary schools are compared.

Graphic 11: Types of expansion hypotactic clauses in secondary schools



If we look at the graph, both Madrid south and Madrid west students mostly use enhancement relations to link their clauses (67.3% and 60.3%, respectively). Enhancement encompasses a large number of subtypes, referring to circumstantial complements: manner, consequence, condition, space and time. The most frequently used subtype is that of cause, as in the given prompt, there are two occasions in which a causal response is elicited. The conditionals are also highly frequent within enhancement relations, specially the second conditional since they reflect on a thought or an idea that is not impossible to happen (“What do you think would happen if Spanish companies were forced to have equal representation of men and women in high-level jobs”). The following clauses exemplify the students’ use of enhancement:

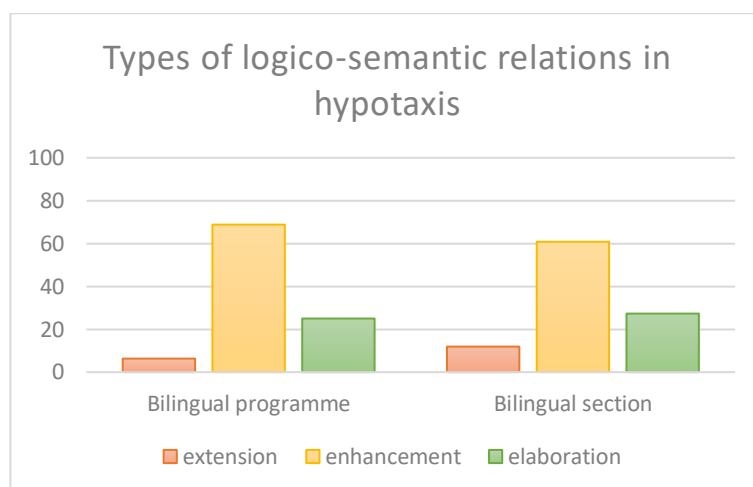
- (32) “Yes, **because** we can learn to manifest our opinions and ideas.”
- (33) “Because **if** the men and women on these companies could be on equal representation in high-level jobs, they all **could** have different ideas.”

It is interesting to note that there is a significant difference in the case of elaboration, with a p-value of 0.0127. In the case of MW students, elaboration accounts for 30,3% of their expansion-hypotactic clauses. As for MS students, their use of elaboration is 21,7%. In terms of absolute value, MW students have established 103 relations of elaboration in contrast to MS students, who have written 67 clauses linked by elaboration. Madrid west students tend to expand the meaning of their previous clause by specifying some aspects of the society, of the women of the past generations, or of the feminist movement. They have proved that they have sufficient ideas about the topic, as well as the means to express them. Instead of producing a simple response to the prompt, there are many cases where they use elaboration to extend the meaning in order to achieve a higher cognitive level:

- (34) “She was the one **who** had to adapt to my grandfather’s conditions.”

The graphic presented below (graphic 14) compares the logical-semantic relations of extension, enhancement and elaboration within the hypotactic relations across bilingual modalities:

Graphic 12: Types of expansion hypotactic clauses across bilingual modalities



Although in the previous graph it was not relevant to highlight the case of extension, since the use of this type of semantic logical relation was quite similar in both groups, it is important to mention it in the comparison across bilingual modalities. The use of hypotactic extension is considerably higher for BS students (11.9%) than for BP students (6.3%). In order to establish hypotactic extension relations, BS learners have made use of more specific and elaborated connectors than *and*, or *but* (paratactic extension connectors):

- (35) “*Women were treated as slaves, **whereas** men could do whatever they want with them.*”

In relation to enhancement, the case of bilingual modalities is similar to the previous comparative case. Both groups mostly use this type of relation as it is elicited through CDFs and it also covers the whole range of circumstantial complements. However, it should be noted that BP students focus more on this type of clause (68.8%) as they make repetitive use of the conjunction *because* in order to link their ideas, and possibly due to a lack of proficiency in formulating a wider variety of sentences. On the other hand, BS students distribute their relations between clauses more evenly, although they give the greatest weight to enhancement (60.8%).

The aim of this part of the analysis was to analyse the type of clause complexes the students have produced as well as the logico-semantic relations they have established between the clauses. In addition, the objective was to compare four groups, MS and MW secondary schools, as well as BP and BS students.

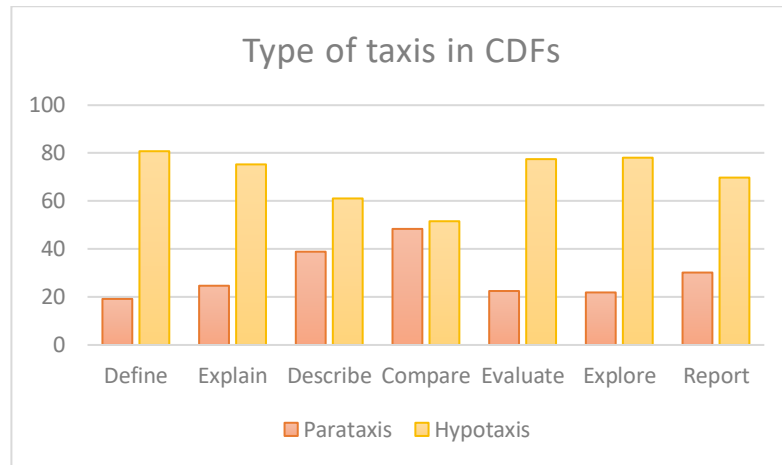
Thus, in order to answer to research question 2(a), we can affirm that grammatical complexity is higher in the case of BS students than in BP students' writings. The conclusions that have been drawn show that the use of hypotaxis is higher in the case of BS students. Contrarily, the number of paratactic clauses in the case of BP students is higher than in the case of BS students. As for the use of logico-semantic relations, BS students have shown a greater command in the English language, as they have been able to establish various relations between clauses. They have not focused just on one relation, but have expanded their meanings by adding information, specifying facts, or contrasting ideas previously stated. That is, they have linked their sentences by recursively combining all the types of logico-semantic relations studied in this paper (projection, extension, enhancement and elaboration) as can be seen in examples 30, 34 or 35, among others.

Regarding research question 2 (b), the results were not significant enough to conclude that the socio-economic status of secondary schools has a direct effect on students' performance. However, students from Madrid south secondary school have shown on some occasions lower English proficiency, as it has been the case in the logico-semantic relation of elaboration. MS students have not implemented this relation as much as the students in Madrid West secondary school, possibly due to their lack of ideas on the topic. Further study would be necessary to draw significant conclusions about this variable.

5.3. Grammar complexity and logico-semantic relations in CDFs

The analysis of grammatical complexity in different CDFs is divided into various sections. First, the following graph shows the percentage of clauses elicited by each CDF that are linked by dependency or interdependency relations.

Graphic 13: Type of taxis elicited by CDFs



In general terms, the seven CDFs elicit more hypotactic sentences than paratactic sentences, so it can be deduced that through questions that stimulate cognitive development, students feel comfortable using more elaborate compound sentences.

The CDF that elicits the highest number of hypotactic sentences is define (80.8%), followed by explore (78.1%). According to Morton (2020), definitions are an example of a typical thinking operation that is asked in examinations in order to achieve educational objectives. Students normally develop patterns of linguistic realisations when they are asked to define a concept. Thus, as students are used to fulfill tasks responding to this CDF, they feel secure to develop their ideas by building more complex sentences. The following extracts illustrate this preference for hypotaxis related to the CDF *define*:

- (36) “When we talk about feminism we refer to the act or demonstration in which most of women raise their voices for their rights.”
- (37) “Feminism is the movement that aims to get equal rights for men and for women.”

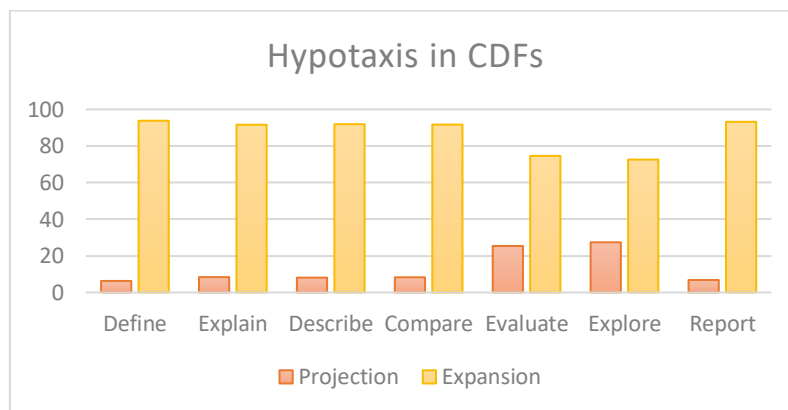
In contrast, the lowest percentage of hypotactic sentences appears in the subcategory *compare*, with only 51.6% of the clauses showing a dependency relation between clauses. By asking students to compare two situations, it is expected that they use discourse markers that introduce paratactic clauses, as students tend to compare two situations using *but* to express contrast. In addition, instead of linking their clauses through connectors or

conjunctions, there are many occasions in which they use comparison markers such as *more... than, compared to, etc.*

- (38) “[...] *she lived in the Soviet Union and she worked as hard or **more than** a man but she didn't get the same merit.*”

The type of parataxis elicited by each CDF is not studied in this part of the analysis because, as we have seen above, most of the paratactic sentences have been established with a relation of expansion, and within this, of extension. Therefore, these data are not significant enough. However, hypotactic sentences are more varied and are the subject of study of this part of the research paper. Thus, the following graph shows which CDF elicits the greatest number of projection and expansion relations.

Graphic 14: Projection and expansion elicited by CDFs



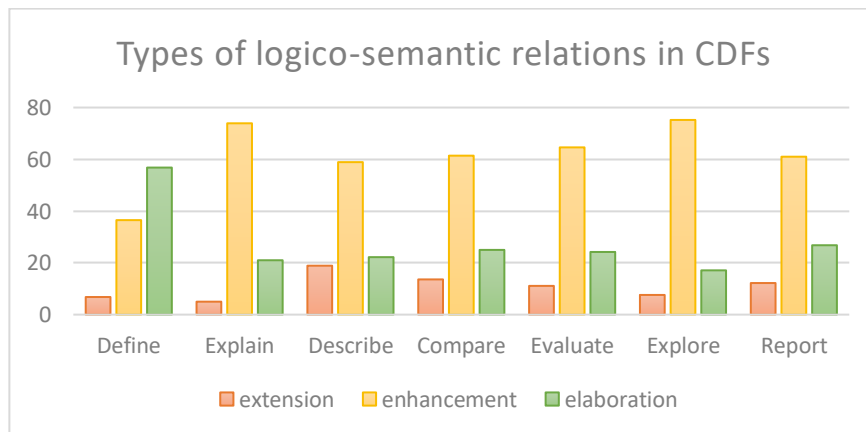
It is worth commenting that projection appears to be very common within the *evaluate* and *explore* function. In the part of the prompt in which the students are asked to evaluate, they had to answer critically whether they think that feminism is benefiting society. In this case, the verb used as the most common resource was *I think*, which served to introduce their opinions and ideas written in reported speech. Thus, although expansion represent 74,6% of the evaluate clauses, 25,4% accounts for projection. As for explore, 27,4% of the clauses are linked by projection. This function asked students to think about the consequences of forcing Spanish companies to have equal representation of men and

women in high-level jobs. In this case the result is similar to what happens with the evaluate function, as students tended to write their answers by using the verbs to think/to believe. Thus, they have used reported speech, that is, projection. Examples (39) and (40) are taken from students' texts in the evaluate and explore function, respectively:

- (39) “I think that thanks to this movement the female collective are now free to decide what they do with their time.”
- (40) “Some people believe that if companies were forced to have equal representation in high-level jobs, the wolrd would became a better place to live in.”

It is of interest in this paper to study the type of logico-semantic relation, within the sentences linked by expansion, elicited by the CDFs. Therefore, graphic 17 shows extension, enhancement and elaboration relations across CDFs

Graphic 15: Expansion logico-semantic relations elicited by CDFs



It is noticeable that there is a general trend for the use of one type of logico-semantic relation. In general terms, *enhancement* is the relation most preferred by students, except in the case of the function *define*. *Enhancenment* is the most predominant relation within *explore* and *explain*, it represents the 75,2% and the 73,9%, respectively.

According to Morton (2020) a good definition should include the term defined, a general word, as well as an expansion. In this case, the students have chosen, in 56,8% of the cases, to expand their definition by specifying or exemplifying their clauses. This CDF is

the one that elicits the greatest number of elaborations, through relative pronouns (*that*, *who*, *which*). The following example taken from the *define* function illustrates how a good definition should be, as it contains the term defined (*feminism*), a general class-word (*movement*) and the expansion (the elaboration clause beginning with “*that*”) (Llinares & Nashaat-Sobhy (2021) and Nashaat-Sobhy & Llinares (2020), as cited in Morton, 2020):

- (41) “*Feminism is the movement that fights for equality in rights and opportunities for men and women.*”

As previously stated, it is remarkable that the functions that elicit enhancement in almost all clauses are *explain* and *explore*. When students are asked to produce a text of a specific genre, in this case an explanation, the language to be used is implicit in the genre. Hence, when students are asked to explain a fact, they usually have to state the causes or consequences. Likewise, the *explore* function encourages learners to imagine a situation, possible or not, and to write about it. There are several subtypes of enhancement relations, so although *explore* and *explain* invite learners to engage in different cognitive activities, they both fall within the framework of hypotactic enhancement. On the one hand, when students respond to *explain* function, they mostly use causal sentences to make a justification, in this case, *why is there a women’s movement today?* On the other hand, in the case of *explore* function, students make use of the conditional to express a possibility. In the study carried out by Llinares & Navarro (2021) conditional sentences were also used to express student’s ideas in relation to the *explore* function. The authors give two possible justifications that can also be given in this paper: firstly, students may have used their linguistic ability to develop a relatively complex grammatical structure as it is the conditional. Moreover, in the prompt itself, the conditional is used as a grammatical structure to introduce the *explore* CDF, so it is also a possibility that the students have directly copied the conditional structure in order to respond to the prompt (Llinares & Navarro, 2021). Examples (42) and (43) illustrate students’ use of conditional clauses:

- (42) “This movement is celebrated **because** the rights of men and women aren’t the same.”
- (43) “[...] but **if** they see that in a period of time women is also seen in other high-level jobs they **would** try to have also representation of women.”

The rest of the CDFs (*describe, compare, evaluate* and *report*), follow the same line. Sentences joined by *enhancement* relation are the most common since students tend to use a lot of circumstantial sentences to express their ideas. *Enhancement* is followed by *elaboration*, as especially BS students specify their sentences by adding characteristics of the feminism movement or of society (“*the feminists that only defend ...*”). It is also remarkable the use of reported speech to express their opinions or ideas (“*she later realised that she didn’t want that for her future*”). Finally, the lowest percentage is for extension relations. Although it is true that sometimes students add information or contrast it through hypotactic relations, they usually produce paratactic clauses by using connectors such as *and* or *but*. The preference for hypotactic connectors may be a signal of high language proficiency. Therefore, students of this age (16 years old) normally choose to use paratactic conjunctions.

To summarize, we can state that the general tendency is for students to write enhancement hypotactic clauses, as circumstantial clauses of time, cause and condition are the logico-semantic relations most elicited by the prompt. It is also worth noting that, although there is a general preference for enhancement, depending on the CDF, students produce different types of relations, as it is the case of the *define* function.

6. Conclusion

This research paper has put its focus on CLIL education contexts. CLIL focuses not only on language teaching, but also on content, so that L2 acquisition and cognitive abilities are integrated and combined. The CLIL approach is of great interest here due to its possibilities within a globalized world, in which communicative and language skills have headed the requirements of the professional sphere.

In order to analyse CLIL contexts, two main aims have been pursued. The first objective was to compare students' language complexity from different linguistic points. In order to do so, 88 texts from 4 groups of students have been analysed. These four groups have been divided taking into account two variables: the socioeconomic status of the city in which the secondary schools were located; and the bilingual modality, whether students attended bilingual programme or bilingual section. In addition, the second aim was to analyse whether the CDFs acted as a bridge between language and content in CLIL contexts. The same data have been analysed to see if depending on the CDF asked, students have produced different grammatical complexity and logico-semantic relations across their clauses.

Through CDFs, meaning-oriented activities can be created to help teachers guide their students in achieving the pedagogical objectives, as their role also involves specifying what students should do with the knowledge. Moreover, if teachers are aware of how to use CDFs to elicit the desired responses, students will be motivated by the fact that they are able to employ English to express and develop their ideas.

Regarding the first research question, the comparison of lexical density and *academicness* between BP and BS students (1a), and MW and MS students (1b), I have come to the conclusion that due to the characteristics of this research, lexical density has proved not to be significant in the determination of students' language proficiency. However, BS students have shown higher *academicness* than BP students. Therefore, it can be concluded from this study that BS students have more developed "academic" English skills to write about a non-specialised topic.

The second research question dealt with grammatical complexity. The very purpose of research question 2(a) was to determine if BS students, as they are taught a greater number of subjects in English, showed higher language proficiency in terms of grammatical complexity and logico-semantic relations between clauses. Based on the results, I have concluded that the hypothesis is confirmed as BS students used more hypotactic clauses

than BP students, who showed more preference for parataxis. Furthermore, BS students have manifested a greater command of the different logico-semantic relations, as the percentages created by UAM CT show that BS students combine and link their sentences through expansion, elaboration and enhancement more frequently than BP students. On the other hand, regarding research question 2(b), the results were not significant enough to conclude that the socio-economic status of secondary schools affects students' performance. This conclusion call the attention for further research on this area, taking into account socioeconomic variables.

Finally, the last research question focused grammatical complexity and logico-semantic relations across CDFs. I have come to the conclusion that CDFs elicit more hypotactic relations as students expand the meanings of their clauses in order to answer to the CDFs given. Students have shown a more frequent use and therefore a higher command of hypotactic enhancement relations, especially on conditional and cause sentences. It is also worth commenting that the *define* function elicits hypotactic elaboration clauses, as students have chosen to to expand their definition of the feminist movement by specifying or exemplifying it.

I would also like to refer to the pedagogical and empirical contributions of this research paper. It has been deducted that trough CDFs students tend to produce hypotactic clauses, as they are guided towards the cognitive task. Furthermore, the analysis of grammar complexity and logico-semantic relations give an overview of students' proficiency. It has been demonstrated that SFL theory is a really useful approach to look at linguistic and cognitive development.

Finally, I would like to mention that this research has some limitations. Although significant results have been extracted, the amount of data studied is relatively small, it would have been interesting to make a longitudinal study based on several tasks, both oral and written, of the same amount of students to see if exposure to bilingual education affects the gradual academic improvement of the student. Moreover, due to reasons of

space, I have just looked up at lexical density and *academicness*, grammar complexity, and grammar complexity elicited by CDFs. It would be interesting to compare if students' responses to CDFs vary across bilingual groups and socioeconomic areas, and also to analyse if CDFs elicit more simplex or complex sentences. Therefore, further research is needed on this area.

Bibliography

- Adrián, M. (2011). An overview of Content and Language Integrated Learning: origins, features and research outcomes. *Huarte de San Juan. Filología y Didáctica de la Lengua*, 11, 93-98.
- Berns, M. (2009). English as lingua franca and English in Europe. *World Englishes*, 28 (2), 192-199 <https://doi.org/10.1111/j.1467-971X.2009.01578.x>
- Busse, V. (2011). CLIL: Content and Language Integrated Learning. *System*, 39, 123-125 [10.1016/j.system.2011.01.001](https://doi.org/10.1016/j.system.2011.01.001).
- Cambridge Online Dictionary. Definition of clause. 2022. <https://dictionary.cambridge.org/dictionary/english/clause>. Date accessed: 22-05-22
- Celce-Murcia, M. (2001). *Teaching English as a Second or foreign Language*. Heinle & Heinle Publishers.
- Comunidad de Madrid. (2016). *Madrid: Comunidad Bilingüe*. <http://www.madrid.org/bvirtual/BVCM016309.pdf>
- Comunidad de Madrid (2019). *Informe datos y cifras de la educación 2019-2020*. <http://www.madrid.org/bvirtual/BVCM050013.pdf>
- Comunidad de Madrid (2022). *Datos ficha Madrid: Boadilla del Monte*. <http://www.madrid.org/desvan/almudena/FichaMunicipal.icm?codFicha=1&codMunZona=0225&tipoFicha=M&codSolapaActiva=5>
- Comunidad de Madrid (2022). *Datos ficha Madrid: Las Rozas*. <http://www.madrid.org/desvan/almudena/FichaMunicipal.icm?codFicha=1&codMunZona=1277&tipoFicha=M&codSolapaActiva=5>

- Comunidad de Madrid (2022). *Datos ficha: Madrid Leganés*.
<http://gestiona.madrid.org/desvan/almudena/FichaMunicipal.icm?codFicha=1&codMunZona=0745&tipoFicha=M&codSolapaActiva=5>
- Consejería de Economía, Empleo y Competitividad. (2019). *Informe del paro registrado por municipios*. Comunidad de Madrid.
<http://www.madrid.org/bvirtual/BVCM007954-2019-12.pdf>
- Council Resolution of 31 March 1995 on improving and diversifying language learning and teaching within the education systems of the European Union. (1995). *Official Journal*, C 207, 1-5. CELEX: [https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:31995Y0812\(01\)](https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:31995Y0812(01))
- Coyle, D. (1999). Theory and planning for effective classrooms: supporting students in content and language integrated learning contexts. *Learning through a foreign language*, 46-62). CILT (Centre for Information on Language Teaching and Research).
- Coyle, D, (2008). CLIL: A pedagogical Approach from the European Perspective. In N. H. Hornberger (Ed.), *Encyclopedia of Language and Education*, 4, 97-112. Springer-Verlag, Berlin. <https://doi.org/10.1007/978-0-387-30424-3>
- Coyle, D. (2008). Content and Language Integrated Learning: Towards a Connected Research Agenda for CLIL Pedagogies. *The International Journal of Bilingual Education and Bilingualism*, 10 (5), 543-562 <https://doi.org/10.2167/beb459.0>
- Coyle, D. (2010). 'Foreword'. In Lasagabaster Herrarte, D. and Ruiz de Zarobe, Y. (eds.) in *CLIL in Spain: Implementation, results and teacher training* (pp. 7 – 9). Newcastle Upon Tyne: Cambridge Scholars.
- Dalton-Puffer, C. (2013). A construct of cognitive discourse functions for conceptualizing content-language integration in CLIL and multilingual education. *European Journal of Applied Linguistics*, 1, 216-253.
<https://doi.org/10.1515/eujal-2013-0011>

- Dewi, R. (2017). Lexical complexity in the introductions of undergraduate students' research articles. *Exposure Journal*, 6(2), 161-172
- Egins, S. (2004). *An Introduction to systemic functional linguistics* (2nd ed.). London: Continuum.
- EF English Proficiency Index (2021). *A Ranking of 112 Countries and Regions by English Skills* <https://www.ef.com/wwen/epi/>
- Eurydice. (2006). *Content and language integrated learning (CLIL) at School in Europe* (Survey). Brussels: Eurydice European Unit.
- Evnskaya, N., & Dalton-Puffer, C. (2020). Cognitive discourse functions in CLIL classrooms: Eliciting and analysing students' oral categorizations in science and history. *International Journal of Bilingual Education and Bilingualism*, 1-20. <https://doi.org/10.1080/13670050.2020.1804824>
- Expansión/Datosmacro.com. *Paro por municipios: Boadilla del Monte, Madrid – (Madrid)*. <https://datosmacro.expansion.com/paro/espana/municipios/madrid/madrid/boadilla-del-monte> Date accessed: 24-05-22
- Expansión/Datosmacro.com. *Paro por municipios: Rozas de Madrid, Las Madrid – (Madrid)*. <https://datosmacro.expansion.com/paro/espana/municipios/madrid/madrid/rozas-de-madrid-las> Date accessed: 24-05-22
- Expansión/Datosmacro.com. *Paro por municipios: Leganés, Madrid*. <https://datosmacro.expansion.com/paro/espana/municipios/madrid/madrid/leganes> Date accessed: 24-05-22
- Gregori-Signes, C., & Clavel-Arroitia, B. (2015). Analysing lexical density and lexical diversity in university students' written discourse. *Procedia-Social and Behavioral Sciences*, 198, 546-556. <https://doi.org/10.1016/j.sbspro.2015.07.477>

- Halliday, M. A. K. (1985). *Spoken and Written Language*. Geelong: Deakin University Press.
- Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014). *An introduction to functional grammar* (3rd ed. revised by Christian M.I.M. Matthiessen. ed.). London New York: Arnold Distributed in the United States of America by Oxford University Press.
- Housen, A., & Kuiken, F. (2009). Complexity, accuracy and fluency in second language acquisition. *Applied Linguistics*, 30(4), 461–473. <https://doi.org/10.1093/applin/amp048>
- INE. (2019). *Atlas de distribución de renta por hogares*. Madrid: Instituto Nacional de Estadística.
- INE (2019). *Tasas de paro por distintos grupos de edad, sexo y comunidad autónoma*. Madrid: Instituto Nacional de Estadística.
- INE. (2020). *Renta neta media por hogar. Base 2013*. Madrid: Instituto Nacional de Estadística.
- Johansson, V. (2008). Lexical diversity and lexical density in speech and writing: a developmental perspective. *Working Papers*, 53, 61-79
- Lasagabaster, D. & Ruiz de Zarobe, Y. (2010). *CLIL in Spain: Implementation, results and teacher training*. Newcastle upon Tyne: Cambridge Scholars Pub.
- Llinares, A. & Dafouz, E. (2010). Content and Language Integrated Programmes in the Madrid Region: Overview and Research Findings. In Lasagabaster Herrarte, D. and Ruiz de Zarobe, Y. (eds.) in *CLIL in Spain: Implementation, results and teacher training* (pp. 95 – 114). Newcastle Upon Tyne: Cambridge Scholars.
- Llinares, A., & Morton, T. (2017). *Applied Linguistics Perspectives on CLIL*. Amsterdam: John Benjamins Publishing Company

- Llinares, A., Morton, T and R. Whittaker. (2012). *The Roles of Language in CLIL*. Cambridge: Cambridge University Press.
- Llinares, A., & Nashaat-Sobhy, N. (2021). What is an ecosystem?: Defining science in primary school CLIL contexts. *Language Teaching for Young Learners*, 3(2), 337-362.
- Llinares, A., & Navarro, A. (2021). Language resources for the expression of content in CLIL: students' use of clause complexes. In Graciano, A., Turchetta, B., Benedetti, F., & Cinganotto, L. (Eds.) *Pedagogical and technological innovations in (and through) Content and Language Integrated Learning* (pp. 58- 79). Cambridge Scholars Publishing
- Marsh, D. (Ed.). (2002). *CLIL/EMILE- The European Dimension: Actions, Trends and Foresight Potential*. University of Jyväskylä.
- Morton, T. (2020). Cognitive Discourse Functions: A Bridge between Content, Literacy and Language for Teaching and Assessment in CLIL. *CLIL Journal of Innovation and Research in Plurilingual and Pluricultural Education*, 3(1), 7-17. <https://doi.org/10.5565/rev/clil.33>
- Nadarajan, S. (2011). The Challenges of Getting L2 Learners to Use Academic Words in their Writings. *Electronic Journal of Foreign Language Teaching*, 8(2), 184-200.
- Nguyen, T. (2013). The Notion of Clause Complex in Systemic Functional Linguistics. *VNU Journal of Foreign Studies*, 29(4), 25-36.
- O'Donnell, M. (2008). Demonstration of the UAM CorpusTool for text and image annotation. In *Proceedings of the ACL-08: HLT Demo Session Companion Volume*, 13 – 16. Columbus, OH: Association for Computational Linguistics. <https://doi.org/10.3115/1564144.1564148>

- Pistorio, M. (2009). Teacher training and competences for effective CLIL teaching in Argentina. *Latin American Journal of Content & Language Integrated Learning*, 2(2), 37-43. <https://doi.org/10.5294/laclil.2009.2.2.14>
- Putra, E. (2020). The Importance of Learning English Nowadays THE IMPORTANCE OF LEARNING ENGLISH NOWADAYS. Sepuluh Nopember Institute of Technology (ITS), Surabaya, Indonesia.
- Qinghua, M & Xinxin, W. (2019). What is Language Complexity? *Macrolinguistics*, 7(2), 1-29. [DOI: 10.26478/ja2019.7.11.1](https://doi.org/10.26478/ja2019.7.11.1)
- Vygotsky, L. S (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press. (Original work published in 1934).
- Whittaker, R. & McCabe, A. (2020). Expressing evaluation across disciplines in primary and secondary CLIL writing: a longitudinal study. *International Journal of Bilingual Education and Bilingualism*, 1-18. <https://doi.org/10.1080/13670050.2020.1798869>

Appendix

Prompt given to students to carry out this study

WOMEN TODAY

Imagine you are participating in an exchange program in the States and classmates and the teacher ask you about the women's movement and the 8th of March in Spain.

Define³ the concept of feminism.

Why⁴ is there a women's movement today?

Describe⁵ what life was like for women in your grandparents' generation and **compare**⁶ it with women's life today

Do you think⁷ the current women's movement in Spain is benefitting society? **Why/why not**?⁸

What do you think⁹ would happen if Spanish companies were forced to have equal representation of men and women in high-level jobs?

Most people in the class have never been to a demonstration. **Tell them what happened**¹⁰ on the 8th of March in Madrid in relation to women's movement.

³ Define (Dalton-Puffer, 2013)

⁴ Explain (Dalton-Puffer, 2013)

⁵ Describe (Dalton-Puffer, 2013)

⁶ Compare (Dalton-Puffer, 2013)

⁷ Evaluate (Dalton-Puffer, 2013)

⁸ Evaluate (Dalton-Puffer, 2013)

⁹ Explore (Dalton-Puffer, 2013)

¹⁰ Report (Dalton-Puffer, 2013)