UCLouvain

FIAL
Un éventail de formations, une multitude d'horizons

Faculté de philosophie, arts et lettres

The acquisition of CLIL and non-CLIL learners' speaking skill in French-speaking Belgium

A corpus-based comparative analysis

Auteur : Alice Van Eesbeek
Promotrice : Prof. Dr. Fanny Meunier
Lecteur : Prof. Dr. Philippe Hiligsmann
Année académique 2019-2020
Master en langues et lettres modernes, orientation germaniques à finalité didactique

Acknowledgments

I would like to express my most sincere gratitude to all the people who have contributed to the realization of this work which is the final step of my academic journey.

First of all, my first thoughts go directly to my master dissertation promoter, Professor Fanny Meunier, who has been of unfailing and devoted support. Her insightful comments allowed me to continue the seriousness of my work until the end.

Secondly, I would also like to thank Professor Philippe Hiligsmann, who believed in my potential and who encouraged me throughout my academic training.

Finally, to all these people dear to my heart who have been present from the beginning. My parents who have always worked hard to give me the chance to go to university as well as my two grandparents Marcel and Emilienne, who from afar have injected me with the desire to perpetuate a family linguistic heritage that was lost along the way.

A special thanks to my greatest source of strength and support, Lise. Lisa, thank you for always standing by my side when spirits failed me.

LIST OF ABBREVIATIONS:

- ARC: Action de Recherche Concertée
- AS-unit: Analysis of Speech unit
- CA: Contrastive Analysis
- CAF: Complexity, Accuracy and Fluency
- CLIL: Content and Language Integrated Learning
- EU: the European Union
- FWB: Fédération Wallonie-Bruxelles
- IL: Interlanguage
- MLU: Mean Lenght of Utterance
- NL: Native language
- L1: first language (mother tongue)
- L2: second language (first foreign language)
- SLA: Second Language Acquisition
- TBKT: task-based approach to language teaching
- TL: Target language
- UG: Universal Grammar

INTRODUCTION

Present-day society is anchored in a world in which multilingualism has become a major asset on social, economic and political levels. Thereby, one might easily grasp the importance given to promoting language learning, more precisely the value of promoting the learning of English as a foreign language. English is indeed one of the five most commonly spoken languages in the world and due to its ubiquity, it has also imposed its status as a lingua franca and represents a symbolic figure of the current age of globalization (Mauranen & Ranta 2009: 1). It is therefore even more significant to highlight the importance of English as a second language on an international stage given its economic and cultural stakes. The growing interest in the development of foreign languages has also proven to be of major interest within the European Union (EU) as it represents and values the concern for adequate economic and politic relations between member countries. In the present case, Belgium and its complex linguistic landscape is at the center of our principal preoccupations as the emphasis of this dissertation is placed on the learning of English within the Belgian educational system. As a starting point, it is pointed out that Belgian students are provided with the opportunity to choose between a traditional language learning education which is included in the school curriculum or *the Content and Language Integrated Learning* (CLIL) system. The first part of this work will be devoted to the identification of the characteristics of the CLIL provision to understand how it functions and to identify what is at stake in comparison to the traditional educational system.

Nevertheless, as we will be brought to discover in the following work, the acquisition of foreign languages is a multidimensional phenomenon. It has been therefore determined that this analysis will focus on the speaking skill's development of the learners in two different groups (CLIL vs traditional foreign language courses). As it shall be explained at greater length in this work, in view of the expansion of the communicative approach within pedagogical studies over the years, oral competence seems to be one of the competencies around which the modern language course is centered as it values the interest of real-time communication interactions. As Sung (2010: 704) argues “the Communicative Language Teaching [CLT] approaches focus on the effectiveness of communication (…) [CLT approaches] and it advocates the idea that second language acquisition develops through communicative tasks, which are class activities that resemble real-world language tasks”. However, even if communicating in real-life context is one of the crucial assets to develop in the modern language classroom, it also represents one of the most complex skills for learners to acquire (Shumin

2002: 204). Due to this apparent difficulty and the thriving need for communication, the productive skills (speaking and writing) are the two skills that current teachers try to develop to the greatest extent possible in order to develop proficiency.

However, although the field that investigates second language acquisition, SLA, has become increasingly important over the past 30 years, several theoretical frameworks have also been developed, leading to sometimes incomplete or contradictory results within this field. As it will be pointed out, few studies have been conducted on the acquisition of learner's speaking skill due to the difficulty of cutting up data into relevant results. The aim of this master's dissertation is to investigate the evolution of the learners' speaking ability within two distinct educational systems, the first being rather oriented towards the development of the communicative approach and the second being dedicated to traditional language learning settings. This present paper is divided into several main sections, consisting of a theoretical part and a practical part. The theoretical part is divided into three main chapters and the practical part is divided into two main chapters.

The first chapter aims to present the *Content and Language Integrated Learning* program, its origins, its international development, its journey into the Belgian educational system and ultimately its beneficial or unfavorable results observed employing previous research on the subject.

The second chapter aims to introduce various theoretical frameworks of Second Language Acquisition that emerged over the years to understand the complexity of this field. Not all of these theoretical frameworks will be used within the practical investigation, only the most relevant ones for our study will be referred to in the analysis. The various theoretical frameworks are mainly intended to develop fundamental notions in order to provide an overview of the field SLA.

The third chapter is intended to introduce key concepts within SLA research, the triad Accuracy, Complexity and Fluency. This chapter traces the emergence and evolution of these measures over time. Then, the psycholinguistic vision of cognition and its direct influence on the three variables are discussed. The next section of the chapter demonstrates the limitations of these variables in order to consider the potential limitations of this current practical analysis. Finally, the chapter concludes by discussing the various challenges of oral proficiency, more

specifically, the challenges of analyzing oral data and the solutions provided for carrying out this analysis.

Once the scientific literature and the potential problems to be encountered have been exposed, a practical analysis is carried out in order to meet the needs of further investigation based on learners' data. First, the methodology implemented in the practical analysis of this MA dissertation is presented. Then, the following chapter discusses the results of the analysis. The analysis focuses on a comparison on the acquisition of the oral proficiency of CLIL and non-CLIL students. Selected variables within accuracy, fluency, and complexity will be used to reveal important aspects of students' proficiency and thus to compare these two groups. Based on the results and the previous scientific literature, conclusions will be drawn about the level of proficiency of the students and it will be possible to discuss the effectiveness of education systems such as CLIL that focus on the development of the communicative approach.

CHAPTER 1: CLIL education and its origins

1.1. Introduction

In recent decades, it has been made possible to combine the learning of a foreign language in the agenda of compulsory course programs. This possibility resulted in the emergence of a number of such type of provision internationally. Nevertheless, this chapter focuses on the appearance of the *Content and Language Integrated Learning* (CLIL) program. To identify its origins, its precursor, the immersive teaching experiment in Canada, will be briefly developed. Based on its huge success in Canada, a great deal of research has been carried out on the subject and programs inspired by this system have emerged in the spirit of the international promotion of multilingualism. At European level, it is the CLIL program that earned its place. In this chapter, the origins of this educational program are discussed as well as its challenges in Europe and more specifically in French-speaking Belgium. Finally, the studies that have been carried out on this subject will also be examined to investigate the kind of learning outcomes this system has to offer.

1.2. Origins of CLIL

Bilingual education appeared in Canada as a result of the experimental immersion experience suggested by English-speaking parents who were concerned that their children speak the two national languages of the country, English and French. Canada possesses a complex linguistic landscape since it was at the governance of European nations during its initial expansion and as a result of that control, the French culture has left an indelible mark on the development of the country (Genesee 1998: 305). In 1969 English and French were recognized as the two national languages but this law did however not apply to the ten provincial governments and only one of these ten provinces is officially declared as bilingual, while the other nine are monolingual. Out of the nine, only Quebec recognized French as its official language, while the other eight opted for English. However, although the linguistic landscape is fragmented, both languages play a decisive role in the country's development and its inhabitants' cultural and economic development (Ibid. 306).

Based on the complex linguistic environment within Canada, a group of English-speaking pupils in St. Lambert were encouraged by their parents to choose French as the language of education in the majority of their courses. Parents' intentions reflected the national

necessity of a good proficiency in French. They assumed that this environment centered around communication and positive cognitive stimulation could help their children become bilingual. These students were entitled to follow an immersive program for three years, from kindergarten to grade 2 and they were taught by native speakers who had to have the necessary qualifications to teach the content and the language (Genesee & Jared 2008: 140). This educational program implies that children received courses in French even before starting to learn to write or to read in their mother tongue, English (Ibid.). This immersive program spread over the years throughout Canada, and in 2008 there were no fewer than 300,000 Canadian students enrolled in this immersion system (Ibid. 141). Due to such enthusiasm, this educational model arose the curiosity of researchers in an effort to determine whether this method performed effectively. Through various studies, it has been shown that these students not only reach an equal level to that of their French-speaking counterparts, but also achieve better results in the foreign language as well as in their mother tongue (Ibid.). Turnbull et al. (2003 as cited in Ibid.) investigated the general English reading achievement of French immersion students and traditional students in grade 3 in Canada. The results show that immersion students are more likely to achieve levels three or four on the English reading test (55%) while their counterparts would be less likely in terms of percentage to achieve the same levels (48%), excluding those who had no English instruction prior to the examination. Results presented by Lapkin et al. (2003, as cited in Ibid.) demonstrate that grade six immersion students outperform the traditional students with a 71% chunk as opposed to a 51% chunk in reading tests. Immersive teaching does not seem to affect first language acquisition, it even seems to activate more advanced cognitive processes that facilitate the general acquisition of a language. With regard to the acquisition of French, a study by Genesee in Montreal (1978, as cited in Ibid. 141) comparing French natives in traditional settings with French immersion students showed that the latter achieved equivalent scores as their French natives' counterparts on standardized reading tests.

1.3. CLIL in Europe

On the basis of the promising results, an interest in the Canadian system has been manifested around the world, especially among countries that were interested in developing the language skills of their inhabitants. The learning of foreign languages has been greatly supported by the European Union in order to foster the prosperity of its construction (Eurydice 2006: 8). As a result, many projects have been launched since the 1950s to accomplish this goal. Some were more significant than others in the creation of this enterprise but all of them

were eventually leading to the promotion of second language acquisition. The European Council summit in Milan in 1985 declared that country members should provide citizens with access to innovative forms of learning to learn at least two other foreign languages within the school curriculum. This summit preceded several expert forums on bilingual education between 1990 and 1996. In 1992, the Maastricht Treaty was a turning point in the promoting of foreign languages as this treaty called for the improvement of linguistic diversity and language teaching methods. "Article 126 states that community action should be aimed at developing the European dimension in education, particularly through the teaching and dissemination of the languages of the Member States" (Marsh 2012: 4). In the course of years, the teaching of foreign languages has become an increasingly important issue in the eyes of the EU. The White Paper that followed this treaty declares that Member States should find innovative techniques to enable citizens to acquire proficiency in three European languages and that secondary schools' courses should be redesigned to be taught in the students' first foreign language, just as in the European schools. That was how the term CLIL came to be coined and how it has become popular since the 1990s (Eurydice 2006: 14). This term is not an equivalent of the Canadian immersive system, but both methods do however share some common characteristics since the latter inspired the appearance of the CLIL program (Ibid. 7). It is an "integrated approach" (Ibid.) that reinforces learning and teaching simultaneously, this approach allows intensive language learning without taking up much of the school's time, since time is devoted to the learning of both language and course content. The term CLIL refers to:

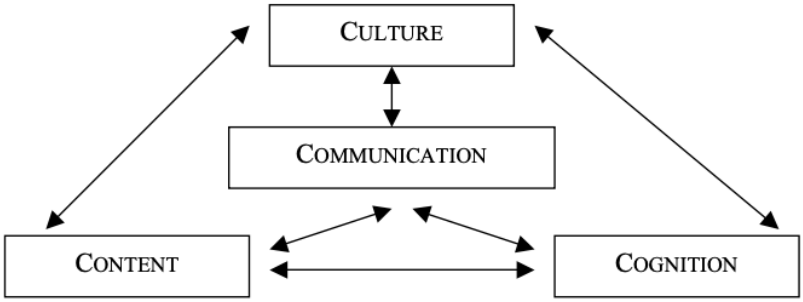
> "all types of provision in which a second language (a foreign, regional or minority language and/or another official state language) is used to teach certain subjects in the curriculum other than languages lessons themselves" (Ibid. 8).

It is therefore important to emphasize that this teaching method has its own methodological and organizational characteristics. One of the advantages offered by the CLIL educational system and presented by Marsh (2012: IV-V) is that this learning method allows the learning of course content and the progression of learners' competences in a more or less natural environment since it takes place in a language bath. Marsh (Ibid. V) argues that:

> "it involves learning environments which have the potential for multi-variant teaching and learning objectives, and experiences. This leads to a synthesis of good practice based on appropriate content (meaningful, new, relevant); incorporation of intercultural understanding (where culture applies to a wide spectrum of forms of diverse interpretation); processing (personalized, peer-driven, and supported); and progression (sequences of learning scaffolded in relation to content and language, and the thinking demands required for progression in each)".

This definition is illustrated in the figure below, which demonstrates the importance of the connections between cognition, content, communication and culture within the CLIL approach. The four levels connect and enrich each other.

Figure 1: A curricular framework for CLIL (Zydatiß 2007: 16 as cited in Dalton-Puffer 2008: 141)



If this system has been so highly valued, it is because this method met the complex requirement of the time that was the development of the communicative aspect of languages. Prior to these days, the learning of languages was based on traditional methods that were used to teach languages such as Greek or Latin. It required a systematic learning of the rules in order to achieve acquisition. However, in the 1960s, the arrival of behaviorism and its vision of language as a structured system, which we will further discuss in the next chapter, led to a shift towards the communicative approach to language learning. Saville-Troike (2003, as cited in Sung 2010: 704) claims that communicative competence encompasses a social as well as a cultural aspect so that learners can translate the forms of a language into meaning. Mehisto (2012, as cited in Marsh 2012: IV), argues that " in the last decade the dual focus on achieving simultaneous content and language learning outcomes has been influenced by multi-disciplinary educational research and dialogue". This vision centered around communicative competence implied the breakthrough of a language learning system such as CLIL.

Beardsmore (1993, as cited in Ibid. III) argues that, as a consequence, the CLIL system made a major entrance on the European market but that each European country could not implement this educational system similarly. Each country has indeed its own national educational policies and its specific social conditions. It would not have been be conceivable to impose a standard CLIL system throughout Europe. Although this program has been adopted broadly across the Union, we must recognize that its distribution has not been uniform

(Eurydice 2006: 51). Nevertheless, the program requires that students must be taught in two different languages which means they receive part of their education in the official language of the community and the other part in the selected foreign language. Depending on the country involved, the foreign language may be another official language of the country, a regional or minority language (Ibid. 8). A total immersion program is therefore not representative of CLIL even though this system varies across the EU and even though the percentage of hours given in CLIL varies from country to country.

As a matter of fact, different characteristics are inherent to this educational type depending on the European countries concerned. First of all, the conditions of access are not standardized (Ibid. 22). Some countries give access to any student who wishes to opt for CLIL education, while in other countries tests are required to gain access. These tests vary according to the countries as well, since they may focus on general knowledge subjects (such as mathematics) in countries such as Slovakia or Bulgaria, just as they may focus on language skills in the target foreign language as a sort of entrance examination (Ibid.). This may be considered discriminatory since some students may be denied access if they do not succeed at the tests. It implies that the EU's desire to promote multilingualism cannot be uniformly respected and implemented.

The main concern that has driven the development of this teaching method is nevertheless of societal nature because the challenges are typically of socio-economic, socio-cultural and linguistic essence but it slightly varies depending on the countries (Ibid. 23). These challenges involve preparing citizens for an internationalized and multicultural world, which will provide them with more professional opportunities. In terms of cultural awareness, this learning method provides an insight into the value system of other cultures and allows to develop awareness towards the world by means of various courses. The student evolves within a linguistic environment that is conducive and stimulating for his effective learning, he uses the language for a specific purpose and his interest is fostered with a view to effective communication. At the level of linguistic added-value, Hiligsmann et al. (2017: 3) refer to the possibility of preserving the learning of a language that is gradually disappearing, such as Basque, within an educational context as minority languages are also promoted through such teaching methods. It also contributes to harmonizing the linguistic landscape of a country with several official languages in countries such as Belgium. It will be observed later on that this lack of linguistic harmonization may have led to profound tensions within the cultural identity

of the inhabitants of the country. Linguistic harmonization eventually leads to a greater scope of employment opportunities and multilingualism is an unprecedented asset in today's society, both nationally and internationally.

According to Eurydice (2006: 55), the interest of the EU in this program is reflected through the added value for the pupils enrolled, due to more advanced cognitive processes' development and linguistic development than it had been previously observed in traditional language courses. This learning environment enables pupils to satisfy the multicultural demand advocated by Europe as we bear in mind that the fundamental challenges of the EU are the desire for unity between different member countries. As stated above, the present study concentrates on the French-speaking part of Belgium and in order to understand the specificities of the program in Belgium, it is first of all necessary to present its historical background.

1.4. CLIL in Belgium

1.4.1. Belgian linguistic and historical overview

Belgium is a country that has been experiencing political and linguistic tensions since it was established as a constitutional independent monarchy in 1830. Within its parliamentary system, bourgeois elites dominated in numbers (Beheydt 1995:50). The issue of language was addressed in the Constitution at the time, which called for "linguistic freedom". However, the reality was quite different and as the bourgeois elites of the time were for the most part French speakers, French thus imposed itself as the obvious choice for the language of the State (Willemyns 2002: 37). Although Dutch speakers made up the largest part of the population, discrimination against this language continued throughout the 19th century and it was not until 1989 that French and Dutch were officially declared as the two national languages (Ibid.). Willemyns (Ibid.) insists that this internal conflict was not only of linguistic nature but also of social and political interest. In 1993, Belgium became a Federal State divided into communities and regions (Chopey-Paquet 2008: 240). The division between the linguistic territories became gradually more pronounced. Nowadays, the country counts three official state languages (French, Dutch and German). The competencies, such as educations or culture, are shared between the three levels of government, namely the Federal State, the Language Communities (Federation Wallonia-Brussels, Dutch-speaking part and German-speaking part) and the Regions (Wallonia, Brussels and Flanders).

This country, in addition to being internally complex due to its division of powers, is an important international hub because Brussels, its capital, is indeed the European capital. In terms of education, although it is the State that finances this area, it is the Language Communities that organize it and pass laws regarding the matter. It implies that the question of education is not managed similarly in the various communities of Belgium.

In order to understand the linguistic division, it is important to study the internal geographical division of Belgium. It is mainly divided in two parts. The most prominent part, which represents more or less 60% of the population, is the Dutch-speaking community in the north of the country while the second part is the French-speaking community which represents more or less 40% of the population and is located in the south of the country. The German-speaking community represents just under 1% of the population (Beheydt 1993: 15). On the grounds of the internal geographic division, the use of the languages mentioned has been established on the principle of territoriality, which claims that the spoken language of specific municipalities depends on the geographical area in which it is located. Brussels Capital Region is however bilingual with French and Dutch being used there (Chopey-Paquet & Amory-Bya 2007: 1). Belgium counts four language regions with Brussels. The value of an education system based on openness to multilingualism is therefore obvious in Belgium and it is within this complex system that the CLIL program landed.

1.4.2. CLIL introduction into the Belgian landscape

The CLIL project was introduced in French-speaking Belgium on an experimental basis in 1989 supported by a non-profit association. An "early English immersion" pilot project started in a school in Liège after obtaining an exclusive ministerial authorization to bypass the language law which called for teaching in the official language of the region (Ibid. 2). It was then legalized in 1998 by a decree allowing immersive foreign language instruction. Only a small proportion of schools opted for this project in the following school year (Eurydice 2006: 15). It was first applied to primary education and it got later extended to secondary education (Van Mensel et al. 2020: 2). Proportionally speaking, Hiligsmann et al. (2017: 7) report that in 2015-2016 a number of 191 primary schools and 100 secondary schools were providing CLIL to over 32,000 students. As aforementioned, there are different types of CLIL within Europe. As far as Belgium is concerned, courses are offered in other languages from pre-primary

onwards and after that, students have the opportunity to pursue in that direction and opt for the CLIL system in primary, lower secondary and upper secondary (Chopey-Paquet & Amory-Bya 2007: 2). These conditions only apply to the French-speaking part of Belgium, since Flanders, the Dutch-speaking part of Belgium, has just recently opted for this type of provision and Dutch-speaking students can only apply to this program in secondary education (Hiligsmann et al. 2017: 4). It will therefore be necessary to continue distinguishing within this work that the different linguistic communities in Belgium do not respect the same regime in terms of education. Our case study will be the French-speaking part of Belgium.

As far as the number of hours in the educational program is concerned, it should vary from a minimum of 25% to a maximum of 75% of the courses given in the target language. In secondary education, this represents approximately 8 to 13 hours of the given classes (Ibid. 6). In addition, there are only very few exceptions as regards the subjects that can be taught in the target language. The only courses that cannot be taught in a foreign language are religion or ethics courses and the French class, which must be given in French (Ibid.). It was initially planned into the legislation to provide flexibility to avoid imposing a single model and that is the reason why the Belgian educational landscape is so much fragmented in the French-speaking part of the country (Van Mensel et al. 2020: 2). Other factors within the Belgian organization such as the wide variation in the level of support from official bodies in the education system and the lack of support for teacher training have resulted in greater flexibility in the implementation of CLIL. Van Mensel et al. (Ibid. 3) therefore argue that there are CLILs in French-speaking Belgium instead of CLIL. In addition to flexibility within the education system, Chopey-Paquet (2008: 242) points out that the linguistic tensions within the Belgian identity might be the reason why the State wishes to promote an immersive type of education in order to promote multilingualism and unify its country.

Although the wish for promotion of multilingualism exists, the Belgian state encounters organizational problems in order to carry out this project successfully. First of all, the recruitment of qualified staff for this specific sort of teaching is not an easy task since these teachers must have a large number of qualifications, namely that "they need to be in possession of a pedagogical degree, they need to have an advanced knowledge of the target language, and they need to have a 'functional' knowledge of French" (Ibid.). It is because such a profile hardly exists today that teachers have started to collaborate with scholars to establish their own educational guidelines (Hiligsmann et al. 2017: 7). In addition to the difficulty of recruiting

suitable teachers, aspiring CLIL candidates generally tend to have different backgrounds and an initial training that was not intended for CLIL teaching. On top of that, these teachers have to adapt and improvise content outside of class hours with very little support from the outside system. There is an increased absence of appropriate teaching materials for this type of teaching (Chopey-Paquet 2008: 242-244; Ibid.).

1.4.3. Underlying problems related to the implementation of the CLIL program

Subsequently, Van Mensel et al. (2020: 3) highlight the underlying elitist aspect of CLIL and assume that this system could be associated with a gap between the achievements and failures of bilingual elites and the non-privileged students. The first group being given an opportunity to be educated in a multilingual environment thanks to determined parents as opposed to the second group to which the same opportunity did not arise (Ibid.). However, Van Mensel et al. (Ibid.) emphasize that the problem is not of economic nature since CLIL education is as free as traditional education in Belgium. Van Mensel et al. (Ibid. 4) describe the phenomenon of the quasi-school market that is prevailing in the Belgian educational system. This term means that FWB schools receive subsidies according to the number of pupils enrolled in their facilities. As a result, schools organize and promote innovative projects or types of teaching to attract a higher number of pupils. CLIL education, which promotes the current plurilingual demand is considered to be one of these innovative types of education and enhances a school's good reputation. The launch of the CLIL project and its intrinsically elitist aspect-initiated schools' competition to be the first to present the project and to collect the most pupils. Van Mensel et al (Ibid. 3) argue that schools offering CLIL attract the so-called, "better" students, because of the "informed consumer" character of educated parents. By wanting to promote intellectual skills such as multilingualism, educated parents wish to provide their children with an environment that promotes intellectual development from an early age and are therefore keen on this type of education. In their study, Van Mensel et al (Ibid. 5) submitted questionnaires on the socio-economic background of parents of CLIL and non-CLIL students. After analyzing the results, they highlighted the significant difference between the groups of students. The study revealed that the parents of CLIL students came from a higher socio-economic background than non-CLIL students (Ibid. 10). Hambye (2009, as cited in Ibid.10-11) also states that French-speaking schools in Belgium that offer CLIL education are particularly attractive to socio-economically affluent families, since they appear to be an environment propitious to the child's good cognitive development, beyond the linguistic added-

value of immersive education. Van Mensel et al (Ibid. 11) point out that this difference is all the more obvious in the case of CLIL teaching of Dutch.

In addition to the elitist aspect of CLIL education, the language studied can also define motivational variables. As mentioned above, Belgium faces deep internal linguistic tensions, which have defined its development throughout its history. The authors support the fact that the attitude of the two linguistic communities in Belgium towards each other is far from positive, contrary to the vision they have of learning English as a foreign language. This means that both language communities are more inclined to learn a foreign language such as English rather than one of the state official languages such as French or Dutch. Dewaele (as cited in De Smet et al. 2018: 52) on the one hand suggests that this negative attitude is correlated with poor learning outcomes in the L2. Chopey-Paquet (2008: 241) acknowledges that "the French-speaking Belgians carry a long-held reputation for being monolingual". Mettewie (2015 as cited in De Smet et al. 2018) on the other hand argues that this conflicting linguistic context is one of the main barriers to the successful acquisition of this L2. Within this research we will not deal with the learning of Dutch among French-speaking learners, but rather with the learning of English. Emotional and motivational factors will therefore most certainly not be subject to societal taboos. Darn (2006: 3) also sells the motivational merit of the CLIL environment for it is interdisciplinary and allows for active and meaningful learning since students can see the benefit of this learning directly.

1.4.4. Research carried out on CLIL outcomes

Finally, in order to conclude this chapter on CLIL teaching, we will report on various studies that have evaluated the impact of this program on the learners' language skills. We start from the assumption that CLIL education is more beneficial than traditional language trajectories.

Just as researchers were concerned for the immersive teaching experiment in Canada, one of the main concerns related to the development of CLIL education is whether it could negatively affect L1 acquisition since part of the disciplinary courses would be devoted to being taught in the target language. While these concerns are legitimate, it has been shown that students in CLIL instruction achieve similar or even higher levels of proficiency than students in traditional settings. Comblain & Rondal (2001, as cited in Hiligsmann et al. 2017:10) have

studied and compared twenty-five children enrolled in a CLIL English curriculum and twenty-five children following a traditional educational pathway. They evaluated their lexical, phonological and morphosyntactic skills in two stages. The pupils were initially tested at the end of kindergarten and secondly after their first year of primary school. According to their results, CLIL students have a better ability to perceive and produce L1 structures considered as simple than their traditional counterparts. From the lexical point of view, the two groups of students developed similarly. Vollmer et al (2006, as cited in Dalton Pluffer 2008: 4) argue that these results are due to the fact that CLIL students need to work more consistently on tasks "showing higher tolerance of frustration, thus acquiring a higher degree of procedural competence in the subject". Subsequently, studies comparing early immersive and later immersive outcomes have shown that the more students are confronted with the target language, the more efficiently they perform. The time of exposure to the language seems to have an impact on the learner's knowledge. It demonstrates the importance of exposing students to the language as much as possible and therefore the importance of input in order to successfully create output in a target language.

Research at the cognitive, linguistic or socio-affective level has been carried out to evaluate the benefits of this system as well. According to Dalton-Puffer (2008: 143), the language domains that are positively impacted by this teaching method are receptive skills, vocabulary, morphology, creativity, risk-taking dimension, fluency, and positive emotional and affective effect. She also discusses the shortcomings of the program, namely the non-impacted areas such as syntax, informal writing, pronunciation and pragmatics. Dalton-Puffer points out that unlike immersive pupils in Canada (Day and Shapson 1996, as cited in Ibid. 144), CLIL pupils are unable to obtain similar abilities as those of native speakers in reading and listening. Dalton-Puffer assumes that CLIL instruction cannot provide as much contact with the language as an immersive environment, but it still contributes to the higher development of receptive skills (Ibid. 145).

In terms of productive competence, Dalton-Puffer claims that at the oral level CLIL students demonstrate greater fluency, quantity and creativity in their productions. It also appears that these students are more inclined to take risks and are therefore more prone to learn more quickly than their traditional counterparts (Naiman 1995, Mewald 2004, Rieder & Hüttner 2007, as cited in Dalton-Puffer Ibid. 144). After a period of time, learners develop less anxiety in situations of spontaneous interaction as they are asked to use the target language in order to

communicate within the classroom. MacIntyre et al. (1998: 547 as cited in Akdemir 2016: 840) refers to the willingness to communicate as "a readiness to enter into discourse at a particular time with a specific person or persons, using a L2" and argues in favor of immersive experiences. He states that this experience is said to have a positive effect on the learners' willingness to communicate. In cross-linguistic research on the influence of third language acquisition, Adrian and Mangado (2015: 186) observed that the CLIL group used the first language significantly less in the classroom than non-CLIL students. This could prove to be due to the interactional quality and communicative environment of CLIL education.

Dalton-Puffer (2008:144) assumes that the main gain from CLIL instruction is vocabulary acquisition. Thanks to context-setting and content classes, the learners' vocabulary expands much more than that of students in traditional courses as they are confronted with a lot more of technical vocabulary. As a matter of fact, overall studies on vocabulary acquisition have shown that the most productively acquired vocabulary is the one that is dealt with explicitly. Matiasek (2005 as cited in Ibid.) points out that the only linguistic variable explicitly addressed in the CLIL environment is vocabulary. Nevertheless, it is important to nuance that it is likely that the acquisition of vocabulary is more significant in the target language since CLIL students are in contact with very technical subjects such as science or geography. Such vocabulary is rarely addressed extensively in a traditional language course, whereas it is obligatory for CLIL students since they have to follow the material in the target language. In the case of this practical study, the vocabulary used by the students will deal with a non-technical lexical field that is accessible to both groups.

Even if several studies suggest that the CLIL educational system is a real launching pad for learners, other studies point the finger at its partial inefficacity. According to Beheydt (2001:15), the CLIL program appears to be more effective for the learners' proficiency in terms of the two receptive skills (listening and reading) while productive skills are inclined to be mastered in a similar manner as students following a traditional second language process. He based his claims on a Canadian study which studied the outcomes of both groups of pupils. With regard to productive skills, Castel et al. (2015 as cited in Hiligsmann et al. 2017: 12) have attempted to demonstrate the positive impact of Dutch CLIL education on the formulation of requests. They found, however, that the results were rather similar to those of their counterparts in the traditional pathway and that the ability of both groups was significantly below the ability of the native speakers. They thereby conclude that the more intensive confrontation with the

language did not positively affect the students' oral competence from a pragmatic point of view. However, CLIL students' abilities in written proficiency do not appear to benefit from significant gains. It could be explained by the fact that this teaching requires hardly any written output from the learners. Indeed, while a language course mainly focuses on the development of the four skills, the CLIL pathway does not offer a language mastery course. This could therefore justify that students cannot progress in a skill they were not asked to practice (Dalton-Puffer 2008: 153).

Based on the results of the research observed on CLIL education, our work will continue with the desire to do further research on one of the productive skills, the speaking skill since it seems that this specific skill is less likely to be developed through CLIL education. In order to define in greater detail what second language acquisition investigates, we will first explore the different frameworks and approaches that have defined the development of this field.

CHAPTER 2: Second Language acquisition

Second language acquisition (SLA) is a research field defined by Selinker & Gass (2008: 1) as "the study of the acquisition of a non-primary language [...], it is the study of how learners create a new language system with only limited exposure […] what is learned of a second language and what is not learned". SLA focuses more precisely on the mechanisms that enable this development, the results of this process and the discrepancies regarding learners' individual differences. The second language which is also often called the target language (TL) is basically any language in the world that one can learn and is also used as a generic term to refer to the learning of a foreign language. It does not necessarily imply that it is the second language learned, but rather a language that is learned in addition to the mother tongue. The wide range of linguistic possibilities of the variety of languages demonstrates the complexity and broad scope that SLA is concerned with. Saville-Troike (2012: 30) claims that success in SLA is therefore far from being obvious and demands strong cognitive efforts in order to be achieved. Succeeding in achieving the same level of complexity in the L2 as in the L1 is indeed very uncommon.

This focus on acquisition has been under investigation for many decades but it started being recognized as a field of its own in the 1960s when scientists formulated for the first time consistent theories and models of SLA. It nevertheless originated from several intertwined sectors such as psychology, linguistics, sociology and other disciplines. One of the main goals pursued by researchers at the time was to apply these theories to improve the language teaching environment and in order to do so, one first had to know how SLA worked. Over the years, SLA took its independence and it nowadays represents a field that is strongly employed to improve language teaching methods, but it does not exclusively aim at improving teaching (Ibid. 36).

Saville-Troike (Ibid. 2) highlights that research which aims at investigating second language acquisition is addressing three questions that are difficult to answer. The fundamental components of these three questions involve learner's knowledge of the language, the techniques that enable him to learn and his likeliness of success. SLA is not a fixed theory because it keeps evolving and because it does not only focus on a particular type of learner's profile. It aims at studying the different levels of acquisition. Authors argue that it would not be possible to find definite answers to these theoretical questions because there are numerous

conflicting factors to be taken in consideration simultaneously. This field is based on theoretical components but also on intrapersonal factors known as learners' individual differences. These individual differences play a major role in the learning process because they influence the learners' learning behaviors and outcomes (Dewaele 2009: 623). Besides, second language acquisition can take place at any moment in one's life. One could be learning a new language right after having learned one's native language, or one could also learn a TL in later stages of life. We must nevertheless distinguish the simultaneous acquisition of several languages in early childhood (which is called simultaneous multilingualism) from sequential multilingualism (which consists in learning other languages after having acquired one's L1). It is indeed possible to be native of several languages. These individual differences have an influence on the results of acquisition (Saville-Troike 2012: 4).

Regarding the complexity of analysis that SLA is concerned with, three main disciplines can delimit and frame the plethora of theories that emerged. These disciplines are linguistics, psychology and sociology (Ibid. 25-26). Divergent or/and complementary frameworks appeared, and it will be demonstrated how important it is to select specific theories while investigating SLA. The combined insights of these fields show how intricate it is to agree on a specific method to investigate data. Researchers have to make decisions and they therefore sometimes have to leave other methods aside, which can lead to inconsistent or incomplete results. Long (2017: 3) states that SLA "has become increasingly fragmented during the past 15 years and is characterized by a multiplicity of theories (broadly construed). The theories differ substantively, and in at least five other ways: source, scope (or domain), content, type, and form". As the present dissertation involves a practical investigation, it will be important to make relevant choices in terms of theory to be as accurate as possible.

The following parts of this chapter will be devoted to different theoretical frameworks that emerged in the course of development of the SLA field. The first approach that preceded the establishment of SLA was *Contrastive Analysis* (CA). This theory still plays a role in SLA studies because some of its components were redesigned to meet the needs of current investigations (Saville-Troike 2012: 40). This analysis falls in the first category that delimits SLA, the linguistic domain.

2.1. Linguistic approaches to SLA

2.1.1. Preludes to SLA: Contrastive analysis

Contrastive analysis (CA) is the systematic comparison of two or more languages, with the aim of describing their similarities and differences. Besides comparing the two languages it carefully studies the structures that could be problematic for learners and it seeks at explaining and solving potential learners' problems. They therefore use one L1 as basis in order to predict the difficulties or/and ease that one alleged learner could ultimately encounter in the L2. One of the pioneers of this theory is Lado, he wrote a book called *Linguistics across Cultures* in 1957. This book explains what CA stands for, it was mainly shaped by behaviorism and structuralism, the frames of reference in the years 1940-1950. It sticks to structuralist linguistics procedures and investigates "the surface forms of both L1 and L2 systems, and [on] describe[ing] and compare[ing] the languages one level at a time" (Ibid. 36). It prioritizes structure before meaning and is syntactically driven. It studies structural aspects such as morphology or syntax. It is nonetheless a systematic process that compares comparable segments one factor at a time. CA pursues the goal of easing the learners' learning process in a target language based on the potential errors that might occur because of his mother tongue. The emergence of this theory is clearly linked to L2 teaching development and improvement. The driving need to compare segments in two different languages allowed scholars to distinguish structures that could or could not be problematic for the learners and therefore allow them to adapt their teaching methods.

Based on behaviorist linguistics, the early stages of CA suggest that the learning success would depend on "habit formation in the process of Stimulus-Response-Reinforcement (SRR)" (Ibid. 26). This principle is based on the fact that one hears a sequence of words, starting from that sequence one repeats it and imitates it. If the process happened successfully, one is reinforced, and learning occurs. Through practice the learner reinforces one knowledge and integrates what is grammatically correct or what is not. Learning a language would be the result of habit formation (Ibid. 36).

Another key factor related to CA is Transfer (Ibid. 37). Transfer is a process that has to do with successful or unsuccessful transposition of structures in the L2 based on the grammatical habits of the L1. Positive transfer stands for learning situations when there exist

similarities between the two languages and where a L1 structure can be directly transposed in the L2. Negative transfer (also called interference), in contrast, occurs when the learner erroneously uses a structure of the L1 in the L2. Lado thought that "the easiest L2 structures (and presumably first acquired) are those which exist in L1 with the same form, meaning, distribution and are thus available for positive transfer" (Ibid. 37).

This theory is therefore based on the fact that the success of one learner will strongly be related to the similarities and differences between his L1 and his L2. Favorable or unfavorable circumstances would be the reasons for learning success or failure. It is true to say that comparing two languages that have more or less the same roots is possible. It is nevertheless not highly relevant to compare two totally different languages which possess two distinct grammatical structures or mechanisms. This approach could be used with any L1 and L2, but the relevance of its results would depend on the roots of these two languages. In addition to that, there is a variety of problems that are at the foundation of SLA theories that cannot be dealt with CA. It does not seek at explaining the issue of the learners' success based on input. In other words, if scholars base their research on comparisons between languages it is not possible to explain how a child masters his mother tongue while he only receives input. We must first start with L1 acquisition studies and investigate where the learner starts his learning process before investigating the issues he will be facing while learning one L2. The learning processes and strategies of his L1 could have an impact on his L2 strategies. Furthermore, problems that are identified by this analysis are not always relevant. It is indeed not always supported by actual learners' problems in actual learning contexts. To sum up, this theory is consistent to prevent learners from making mistakes if the two languages are easily interconnectable. According to Xie & Jiang (2007: 10) "the principal barrier [of this theory] to second language acquisition is the interference of the first language system with the second language system". This approach was a predecessor of SLA.

2.1.2. First theories of SLA

Saville-Troike (2012: 26) argues that linguistic research within SLA has been led from two different angles over the course of years. The first is the internal linguistic approach which was led by Noam Chomsky (1957). Chomsky claimed that the behaviorist approach was too limited because it did not consider the creative aspects of language learning processes. He therefore formulated his first theory on the innate language capacity and developed it through

the theory of "Principles and Parameters" and the "Minimalist Theory". The theory of Universal Grammar (UG) combines these theories developed by Chomsky. UG is an elementary linguistic principle which directly concerns first and to a certain extent second language acquisition. This theory has allowed scientists to find a lead to explain how children are able to master their mother tongue based only on what they hear, the input. We will therefore start our reflection with this theory because it allows us to consider the likely innate processes that influence and enable language acquisition.

2.1.2.1.UG

As previously stated, this theory arose because researchers were trying to explain why children manage to acquire a native level in their mother tongue and exclude ungrammaticality from their language system by receiving input. There are indeed complex grammatical structures that natives assimilate without having received theoretical explanations for their uses. The expression "poverty of the stimulus" is used as input only cannot provide an adequate explanation for the relevance of use of certain specific grammatical structures. White (1989, as cited in Selinker and Gass 2008: 162) exemplified this statement with the informal contraction form «wanna» of «want to» in English that cannot be interchanged in every sentence containing "want to". There are a number of cases where this contracted form will not be a proper fit. For instance: Who do you want to feed the dog? *Who do you wanna feed the dog? The fact that the contracted form cannot be used in this context, has to be explained to a learner of English because there are no explanations provided in the environment of the utterance (Ibid.). Chomsky argued that language-learning environment does not provide sufficient information and made the assumption that learners' competence relies on innate universal knowledge, the language faculty. According to this theory, language would represent "a set of abstract principles that characterize core grammars of all-natural languages" and "UG is postulated as an innate language facility that limits the extent to which languages can vary" (Ibid. 161). Saville-Troike (2012: 47) adds that "knowledge of principles and parameters is postulated to be innate; children are assumed to be able to interpret and unconsciously analyze the input they receive and construct the appropriate L1 grammar".

Selinker & Gass (2008: 160-163) explain that Chomsky extended his theory and explained what this innate capacity stood for and developed the theory of principles and parameters. This notion of principles represents "properties of all languages in the world", they

are invariable because "all languages have them" (Ibid.160). Selinker & Gass argue that "the basic concept behind this theory is that linguistic principles operate on syntactic (structural) units" (Ibid.168). This means that the relevance of this theory is based on considering language as a structured system based on the dependency of its segments. The principle of UG is to consider language knowledge as something that is fixed and structured. According to these authors, what makes language so different from any other type of knowledge is the "notion of structure dependency" (Ibid.).

This theory serves as fundamental basis because it studies the "perspective of learnability" (Ibid. 160) of first language acquisition and by extension the learnability of second language acquisition. There are nonetheless some limitations to this theory because if innate language faculty compartmentalizes the grammar possibilities, it would imply that "the task for learning is greatly reduced if one is equipped with an innate mechanism that constrains possible grammar formation" (Ibid. 161). Therefore, the nature of the linguistic knowledge with which the learners begin the second language acquisition process should be examined to link UG to SLA. In order to make a link between the two, hypotheses relating the access that humans would potentially have to UG in the process of second language acquisition will be discussed as well as the notion of transfer.

2.1.2.2. The Fundamental Difference Hypothesis

Many scholars argue that learning a native language is fundamentally different from learning a second language and that the acquisition of these two cannot be compared. This theory is called *The Fundamental Difference Hypothesis*. This claim is based on a wide range of arguments. Firstly, the level of linguistic knowledge of adults learning a L2 and that of native children appears to be different in many ways. The level of acquisition is most likely not equivalent because adults have a tendency to fossilize while children learn their native language without falling into a fossilization process, at least during the first years of their lives. Fossilization represents a moment in the learning process where the language competences cannot be taken any further and where the level of linguistic knowledge cannot be extended.

Secondly, children learn in a native environment the form and meaning of words as well as their social connotations while adults have to learn that some specific words are used in definite contexts. This learning does not come naturally and neither does it come in a native environment when it occurs in the classroom.

Thirdly, scholars assume that the learnability of the second language relies on the first language. For instance, an English native speaker will learn languages with Germanic origins more easily than, say, Korean because these languages belong to different language families that differ greatly. Children however will face the same difficulty while learning English or Korean as a first language because they do not rely on a previous internal grammatical system to construct new linguistic knowledge (Ibid. 164-165).

Lastly, motivational factors are also divergent for the two groups. Humans are not gifted with the same motivation when they learn a foreign language. This aspect plays a major role in the progression of the proficiency in the second language. This motivational factor does not seem to apply to children because they must learn, it does not rely on their motivation and own desires in order to be successfully achieved. To sum up, Selinker and Gass (Ibid. 165) argue that the Fundamental Difference Hypothesis assumes that adults do not access UG in order to learn a second language. The scholars who support this theory state that "what they know of language universals is constructed through their NL" (Ibid.).

However, as it has been previously explained, there are a number of other theories that claim that adults can access UG to learn a second language. They will not be discussed thoroughly in the following part of this work, but an important shared perspective between these theories is that the learner possesses his L1 grammar and an access to UG when he starts to learn a foreign language. For some scholars, the L1 serves as basis while others consider that the learning process is UG-based. The type of access to UG will be determinant in terms of transfer issues according to these researchers, but these theories are not essential in this specific practical framework. We will therefore say that there exists different point of views, but we will not discuss them any further.

2.1.2.3. Functional/External approach

As seen above, Chomsky's approach focused on language structures and his vision of language was systematic and based on an unconscious system that guides learning. The difference with the external approach has to do with the fact that language is considered as a communication system. The content and information of the utterance are what matters most within this other view. The framework that fits best in this new category is the theory of Functionalism. It emerged in the early 20s in the Prague school of Eastern Europe (Saville Troike 2012: 56). Scholars inserted a new notion within the meaning of the word "function". The function could be either structural or pragmatic (Ibid.). The pragmatic function represents a way to carry information and to carry emotions through the system of communication. It basically consists in "what the use of language can accomplish" (Ibid.). The structural function is meant to structure discourse.

A number of functionalist theories emerged as well and these theories do not always share the exact same features but what they do have in common is that researches are oriented towards the fact that "the purpose of language is communication, and that development of linguistic knowledge (in L1 or L2) requires communicative use" (Ibid.). The different theories based on Functionalism will not be developed, but it is nevertheless essential to see the shift in the field. In fact, it is moving from a structural vision of language based on an internal grammatical system to a vision that leans towards a communication system that allows humans to control behavior or express emotions.

2.1.2.4. Error Analysis

This concept is based quite explicitly on the analysis of errors. There is a distinction between Contrastive Analysis and this new approach because "the comparison made is between errors a learner makes in producing the TL and the TL form itself" (Selinker & Gass 2008: 102) whereas CA started from a comparison between two languages. They however both start their investigation from learners' data. This new theory is based on the active participation of the learner and its environment in order to understand the level of knowledge he possesses in the TL. It is seen as a creative and evolving process. Compared to previous theories, it can be observed that the learner is seen as one of the main actors in his learning progress. The type of errors detected in leaners' productions is used to categorize his evolving process and provides an insight into learner's language system. Corder (1967) wrote one of the most significant

papers on this issue. He argues that errors should not be considered as "bad habits" by applied linguists but "as sources of insight into their learning processes" (Saville-Troike 2012: 41). The knowledge acquired and the learning development system of the learner are what matters most in this process. The strategies that the learner uses are brought into question to see what type of language he is capable of using. This approach is based on the fact that when learning a new language, one is exploring a new system and not transposing native habits to the TL. This theory shows that the learner does not lack imagination in his learning process but that he encounters specific issues while trying to figure out a new system in the second language. Corder (1967) made the distinction between the terms "errors" and "mistakes" because mistakes are, according to him, "generally (a) one-time-only events" while errors are something "systematic" (Selinker & Gass 2008: 102). The mistakes are noticeable and can be corrected by the learner himself. The learner is not capable of recognizing what he produces as an error. This means that in his mind, what he produces is not considered as an error and only teachers or natives could correct him. Selinker & Gass (Ibid. 103) say "errors are only errors with reference to some external norm (in this case the TL)".

Furthermore, this approach is oriented towards language teaching improvement. Selinker & Gass (Ibid.) argue that "a great deal of the work on error analysis was carried out within the context of the classroom". The aim and the final step of its methodology is to remediate to learners' problems.

As aforementioned, there are also issues regarding Error Analysis. First of all, it relies on errors to the cost of other information that could be interesting in the learners' productions. It is indeed a fact that "one needs to consider nonerrors as well as errors to get the entire picture of a learner's linguistic behavior" (Ibid. 104). A second issue of this theory has to do with the categorization of errors. It is indeed not always an easy task to classify what is considered as an error and what it is not. There is a distinction between the structure that was first intended and the result of the attempt. Finally, the nonexistence of certain errors in learner's speech might be due to the fact that learners avoid complex structures because they are uncomfortable with them (Saville-Troike 2012: 43).

2.1.2.5. Interlanguage

Corder, Selinker and other researchers developed the theory of Interlanguage (IL) in the years 1970 but Selinker was nonetheless the one who coined this term. Interlanguage is defined as "a continuum between the first language and the target language along which all learners traverse" (Rustipa 2011: 19). Learner's role is significant in this view because interlanguage represents a system that emerged between the L1 and L2. We can observe that as the years went by, the development of SLA became more oriented towards the learners' role in the process and the communicative value of languages.

However, Song (2018: 17) argues that at the time "they still used error analysis in studying second language production, but their goal was to [...] figure out the second language learner's cognitive strategies for grasping the TL system". Interlanguage indeed legitimizes learner's new language system, which differs greatly from the L1 and the L2 and which is a sort of L3. It is a creative process influenced by the learner's environment. It is not a combination of NL and TL elements but knowledge which emerges outside the L1 and L2 (Selinker & Gass 2008: 14). On the basis of this theory, the acquisition of a second language should not be considered as a result of transfer, but it is nonetheless correct to say that the native language influences the competences in the second language (Song 2018: 17). The main features of this process are related to the learner's creative mind in understanding new characteristics of the target language and the legitimization of the creative system of acquisition rather than considering it as a distorted system filled with errors (Xie & Jiang 2007: 11). "Interlanguage is an outcome of second language learner's cognitive learning and proves the learner's competence in the target language" (Song 2018: 17). Song (Ibid.10) argues that "with the introduction of Interlanguage, scholarly attention moved to language learner's cognitive mind. This approach explains second language learning process with the human internal mechanism of language learning".

Another term that is strongly related to this approach is fossilization, namely the "probability that the[y] learners will cease their IL development in some respects before they reach target languages norms, in spite of continuing L2 input and passage time" (Saville Troike 2012: 41). As a conclusion, it can be observed that the cognitive creativity of the learner has become a significant factor in SLA studies with this new notion. The principle of fossilization allows to discuss the limits of these processes for adults.

2.2. Psychological approach to SLA

The second field that frames SLA theories is psychology. The relationship between the brain and language learning, learners' individual differences as well as learning processes are among the topics that are addressed within the psychological approach of SLA. Some theories offered by neurolinguistics on the link between the brain and language acquisition will be briefly presented. We will then focus on research on learners' individual differences which rely on a humanistic approach (Ibid.72).

2.2.1. Brain

One of the most important discoveries about the brain and its relationship to language can be traced back to the 19th century (Ibid.). Broca identified a part of the brain, the left frontal lobe, as the part that manages the ability to speak. He claims that if this part were to be damaged it could result in much more language-loss than if the right part were to be. Shortly afterwards, Wernicke identified a part in the cortex that manages audio input as well as language processing. Over the years, further research has shown that language ability is not only located in the left hemisphere of the brain (Ibid.). However, "core linguistic processes" are typically located in the left side of the brain. A key notion related to the brain and language acquisition is the plasticity of the brain (Ibid.89). This term defines the ability of the brain to change its connections so that certain functions can take over in the event of damage to a part of it. Lenneberg (1967) argues that humans have a limited number of years to learn to speak their L1 natively before starting to suffer from brain damage in the language areas, resulting in less plasticity of the brain. This theory is called the Critical Period Hypothesis (Ibid. 73).

Within neurolinguistics studies, other important research has been carried out on the effect of bilingual education on the brain. Dewaele (2009: 623) states "brains, like fingerprints, are unique" in a wish to explain that the different connections within the brain evolve during one's existence and because one's knowledge is affected by one's doings. Osterhout et al. (2008 as cited in Marsh 2012: 398) advocate the benefits of L2 learning in classrooms settings. According to three research projects, the L2 teaching environment can lead to electrical changes in the brain and can restructure its connections. Marsh (Ibid. 66) points out that researchers had made the assumption at the time that second language acquisition was most likely to have a positive influence on the brain's structure but only if the learner had a high level of proficiency

in different languages. This belief has been challenged ever since and indirectly promotes the potential benefits of immersive learning for learners and Coggins Kennedy & Armstrong (as cited in Ibid. 333) claim that bilingual learning may have a powerful impact on brain processes. We therefore return briefly here to the interest of the Content and Language Integrated approach. As it was discussed in the first chapter, this type of teaching has positive effects on the development of human cognition and these findings confirm this assumption. The impact of immersive instruction would go beyond language skills and it would also promote other cognitive skills. Mechelli et al. (2004 as cited in Dewaele 2009: 623-624) observed the expansion of gray matter density in the brain of bilingual individuals when compared with the brain of monolinguals. Marsh (2012: 398) claims that “this has implications for not only recognizing the value of partial language competences, but also for understanding why certain approaches such as CLIL appear to lead to positive learning outcomes”.

2.2.2. Learner individual differences.

Learners’ individual differences influence the way an individual learns languages. It has been shown that even if two learners are equally exposed to a foreign language, they will not perform similarly in terms of TL skills. This is due to these subjective factors and this topic is multidimensional as well. First, the factor of ability. Gardner (2006, as cited in Dewaele 2009: 625) referred to this term as "both intelligence and language aptitude" and states that the higher the level of ability, the more successful the learners will be. The vision of the relationship between language learning ability and intelligence has evolved over the years. Dörnyei (2003, as cited in Ibid.) has even come to question whether there is a link between the two or whether intelligence should be taken separately from language learning ability. Dewaele (Ibid. 626) demonstrated in a study that in high school, the students' scores in their first language were linked with scores in their second languages and argues that "these higher scores in language classes could be linked to cognitive or social factors, or a combination of both". Miyake and Friedman (1998 as cited in Ibid.) said that the central element of language fluency was related to the working memory as we will discuss in the final chapter of this work.

Other identity factors such as age, gender, motivation, cognitive style, personality and learning strategies are significant factors in studies of learners’ individual differences. From the point of view of age, we can first of all reconsider the Critical Period theory, which is based on the principle that beyond a certain age a learner will lose its brain plasticity in a progressive

manner, while this ability allows him to master his native language. This theory would for instance explain why learners keep track of influences from their L1 in their L2 pronunciation systems. Pronunciation appears to be influenced by the lack of plasticity over the years. Secondly, what differentiates adult learners from child learners are the brain skills. Due to their young age, children are not equipped with an analytical mind as it will develop over the course of time. This absence of analytical skills enables them to adapt more easily to informal learning contexts. They would also appear to have a weaker sense of belonging to their speaking community than adults. This ability to adapt is less common for adults and it represents a disadvantage in the cognitive process of foreign language learning.

Another personal characteristic that has been widely studied in SLA is motivation. Saville-Troike (2012: 92) distinguishes emotional motivation from instrumental motivation. Instrumental motivation is the purely utilitarian view that the learner perceives of the foreign language. It can for instance be linked to instrumentalized added-value as it represents a professional challenge for the learner. Subsequently, the affective factor is the desire to join the speaking community of that specific language, the desire to join this group by sharing language knowledge. As far as our future practical study is concerned, it is important to underline that students are also subject to motivation when they learn a foreign language. However, the compulsory school context leads to a multitude of attitudes towards learning. It is therefore interesting to investigate whether students following a traditional pathway and students following a CLIL program share the same motivation.

One of the last significant personal factors we will discuss is personality. Personality traits such as anxiety, self-confidence, empathy, fear of risk influence how learners learn. These factors are indeed all the more emphasized in a school context since learning takes place in a group. These personality traits are therefore potentially challenged because of social pressure "anxiety correlates negatively with measures of L2 proficiency including grades awarded in foreign language classes" (Ibid. 96). However, if anxiety plays an important role in the development of the learner's interlanguage, it is likely that the CLIL teaching environment promotes the aspect of confidence and familiarity which is apparently necessary for the learner to dare and learn more easily.

In addition to that, MacIntyre, Clément, Dörnyei and Noels (1998 as cited in Dewaele 2009: 630) have proven that personal development flourishes through social contact and experiences. It is therefore obvious that environments that promote the safe space for language learning are supposed to be more efficient for the learners' development. Consequently, it can be observed that, depending on the country, the interest in learning a second language differs from country to country. Dewaele (Ibid. 63) mentions countries such as Belgium, China and Poland to evoke student awareness, without implying the motivation to achieve the given goal, of acquiring knowledge in another language, unlike countries such as the United States or the United Kingdom where international pressure does not require the learning of a foreign language. It would appear that the parents of French- and Dutch-speaking students in Belgium would be very supportive of bilingual education in the interest of promoting their economic future (Beheydt 1993: 17). Nevertheless, as mentioned in the first chapter, this factor in Belgium is also influenced by the foreign language concerned.

2.3. Conclusion

On the basis of the previously stated theories, some important concepts can be identified for the further development of this work. Firstly, the notion of transfer can be highlighted. The practical data includes conversations between Francophone learners studying English as a second language. It is possible that in the analysis of the data, a pattern of transfer will be found. Indeed, since oral proficiency is probably one of the most difficult skills to master, it is possible that students might rely on L1 structures to facilitate interaction.

The vision of language as a structured system can also have an impact on learners' knowledge construction. In fact, we can talk about combinations of words frequently used within the language, chunks. It is likely that as a result of guided learning in the classroom, students are more likely to use structures they have learned and have assimilated based on habits formation. However, it will not be possible to focus on chunk analysis in the quantitative study of the data as the practical investigation will primarily focus on learners' proficiency.

The theory about learners' individual differences is also relevant for this analysis since it will focus on students who have followed a traditional educational path and on students who have chosen to follow CLIL instruction. It is possible that in terms of motivational factors,

students who have chosen the CLIL pathway will be more uniformly motivated than traditional students and that they will therefore outperform the latter.

Furthermore, the theory of Error Analysis appears to be relevant because it starts from learners' data and it considers them as fundamental actors in the learning process. It is based on the principle that errors found in the learner's language system enable us to identify his gradual acquisition process. However, this analysis is based solely on errors in the TL compared to the TL norms and it does not start from the native language nor does it study the non-errors. It is not yet defined if the practical investigation will deal with transfer issues. However, it can be noted that all learners concerned are native French speakers. It is therefore likely that this factor will affect our results and the research choices.

The theory of Interlanguage is relevant as well since it considers the learner's language system as a system in its own right that deviates from the L1 and L2 systems. However, it should be pointed out that, from this point of view, the native language still influences the learning process but is not the basic element of it. It is through this term that the status of learners' knowledge will be considered. In order to deepen the research in proficiency measures, the variables complexity, accuracy and fluency are studied in the following chapter as well as the psycholinguistic vision of cognition.

CHAPTER 3: Complexity, Accuracy and Fluency

3.1. Evolution

The triad *Complexity, Accuracy and Fluency* (CAF) has been making its way in SLA research because it provides the means to assess learner's proficiency and performance skills in-depth. Proficiency is typically defined in its broadest sense as "a person's overall competence and ability to perform in L2" (Thomas 1994, as cited in Bulté & Housen 2015: 50). This concept is neither unitary nor easily determined and Bulté & Roothooft (2020: 2) argue that the definition provided by researchers is generally too vague within CAF studies.

Nevertheless, the growing interest for this triad started in the 1970s. Researchers at the time were trying to turn to more objective and quantitative measures that came from L1 studies to measure L2 development. They therefore decided to investigate grammatical complexity and accuracy. Shortly after that, Brumfit (1979), while doing his research on L2 pedagogy and by extent L2 proficiency, distinguished the term *accurate* and *fluent*. Fluency was from then on, the variable that dealt with the spontaneous nature of learner's oral productions while accuracy was to be investigated through the linguistic and grammatical forms used in L2 productions. At the time, the "complexity" dimension was not to be compared in regard to these terms as it was introduced in the model by Skehan in the 1990s. This last incoming variable turns out to be the most complex and the least transparent to investigate because of the multifaceted nature of its components, which originate on the one hand from objective measures and on the other hand from subjective data such as learners' individual differences (Housen, Kuiken & Vedder 2012: 2-4). Overall, complexity is defined as "the ability to use a wide and varied range of sophisticated structures and vocabulary in the L2", accuracy as "the ability to produce target-like and error-free language" and fluency as "the ability to produce the L2 with native-like rapidity, pausing, hesitation, or reformulation" (Ibid. 2). Over the years, interest in the triad has developed greatly. While L2 studies began to focus on the cognitive mechanisms involved in language acquisition, these variables were to be considered as "the primary epiphenomena of the psycholinguistic processes and mechanisms underlying the acquisition, representation and processing of L2 systems"(Ibid.).

At the same time in the 1970s, as already mentioned, studies in pedagogy shifted with the communicative approach. This approach favors the use of speech as the transmission of meaning and information within linguistic messages. This new vision shaped the pedagogical

methods at the time, and it created new teaching activities centered around it. It was no longer sufficient to focus solely on the study of language structures because it had become essential to realize the communicative capacity of language productions. At the heart of pedagogy was henceforth an interest in the "communicative activity" (Skehan 2003: 1). This term was soon replaced by the word "task" and led to a task-based approach to language teaching (TBLT). The task represents "an activity which requires learners to use language, with emphasis on meaning, to attain an objective" (Ibid.3). This change of vision influenced SLA studies as well, as the shift towards a communicative approach falls within the scope of psycholinguistics research. It means that it studies the processes involved in the learner's language learning process such as the role of memory and type of knowledge accessible to the learner while he produces speech. This change lead to a more obvious connection in regard to CAF as it conceptualizes the learner's performance in a cognitive way (Skehan & Bui 2018: 1).

The relevance of CAF as an independent and primary variable in SLA studies has since then been demonstrated both theoretically and empirically. Empirically, it has been shown that these variables have been recognized as separate domains of L2 performance through factor analysis. It implies that these three components are necessary if researchers wish to report on the status of the learner's performance and competence extensively. At the theoretical level, the three variables have proven their contribution in the change of the state of knowledge in the TL system (Housen, Kuiken & Vedder 2012: 3).

In terms of cognitive changes within the L2 system, the assumption of Housen, Kuiken &Vedder (Ibid. 5) is that the complexity of the learner's language system is determined by his "declarative linguistic interlanguage knowledge". It represents the grammatical rules or the formulaic skills which have been integrated at a given moment under the constraint of phenomena such as UG or transfer for instance. The change of state from implicit to explicit knowledge, which is in other words the assimilation of knowledge, represents the way in which the L2 system changes. The stages of this process would be divided in three phases. At first the learner is confronted with the "internalization of new L2 elements" and then he faces the "modification of L2 knowledge" through contact with new language features (Ibid. 3). These two stages represent the numerous encounters with new language elements. The process of building and modifying knowledge on the basis of the new information leads to the complexification of integrated basic knowledge. It is at this stage that learners regulate their set of skills. It implies that it is a moment in their learning process where the deviances that could

be found within their interlanguage are modified. This phase increases the complexity of learners' knowledge which allows them to move on to the last stage, namely the "consolidation and proceduralisation of L2 knowledge" (Ibid.). As aforementioned, these phases occur while shaping implicit knowledge, namely it is the process of acquisition of L2 competences. However, as mentioned above, it is possible that during its acquisition the learner may find himself in a fossilization phase. This would imply that he unconsciously could not go beyond in terms of knowledge. The three triad variables play a role in this process since they are the variables that assess the development of knowledge in terms of proficiency.

Although they have proven their legitimacy within SLA studies, the relationship that these three components share is nonetheless subject to much debate. It will be observed later on that according to some researchers the consolidation of certain abilities can lead to the loss of others. Other researchers, such as Robinson, argue that one task can lead to the simultaneous improvement of several proficiency variables in the learner's interlanguage. The difficulty of this investigation is related to the unconscious cognitive processes intrinsically linked to these three variables.

In view of the definitions provided, it is noted that these components interact with each other and that they are essential to study the learner's skills development. However, the complexity of their internal features makes the interaction between them relatively difficult to study. In an attempt to evaluate how this triad can best be integrated into the practical study of this master's dissertation, the three components will be defined as clearly as possible. It is therefore necessary to look for tools that can be best suited to this research. First of all, the psycholinguistic vision of CAF is presented. Subsequently, the major problems of the three variables will be listed. On the basis of these observations, the variables that will be used in the practical analysis of this work will be determined.

3.2. Psycholinguistic vision of CAF

One of the fundamental elements of the psycholinguistic perspective is cognition which

> "refers to the processes of human learning in real time, understood in terms of human information processing, whereby information is encoded, recoded, decoded, i.e. translated into other formats, as well as chunked, elaborated upon, stored, retrieved and used—constructively as well as creatively" (Barsalou 2009, as cited in Drozdzial-Szelest & Pawlak 2014: 6).

This definition reflects a fairly procedural way of looking at languages, encoded in major principles such as memory, planning and anticipation processes (Ibid. 10). The psycholinguistic vision of these variables is the one that will be adopted in the formulation of the following hypotheses.

Within psycholinguistic studies, the place of mental mechanisms is highly important. According to this vision, it is assumed that there are three types of mental representation. The three must be mobilized by the learner in order to learn efficiently. The first principle claims that the learner must "acquire an appropriate mental representation for linguistic competence" (Towell 2012: 49). This mental representation is associated with an abstract process in order to achieve the use of linguistic phenomena such as morphology or syntax. It functions via internal mechanisms such as UG which is a set of abstract techniques that allows the realization of syntactic phenomena such as word order as discussed in the second chapter. These procedures are linked to each other and each language has its own selection of processes.

The second mental representation consists in "being able to represent learned linguistic knowledge"(Ibid.). It consists of knowledge as found in dictionaries, i.e. the specification of syntactic, stylistic or even pure pragmatic rules. This knowledge is the result of learning through explanations, it is assimilated based on theoretical information. Finally, the third is "to build a suitable mental representation for the procedures which enable the processing of language in real time"(Ibid.). These procedures are based on a condition/action relationship that the learner uses to build knowledge. The effectiveness of these procedures will increase as they will be used. It implies that the learner might need less time to access them as the condition/action mechanisms become more automatic with time and use.

After the mental representations, another important point in psycholinguistics is the different types of learning. According to these theories the triggered, explicit and procedural learning are to be considered when one refers to learning (Ibid.50). In a few words, triggered learning is based on syntactic skills acquired on the basis of an unconscious process. It consists in the ability of learners to decode the clues present in the language structures they encounter and to set parameters according to what they have perceived. The learner uses unconscious mechanisms such as UG to take advantage of this unconscious decoding of language data and assimilate what he decoded (Ibid.).

Explicit learning is conscious, and the information is processed quickly but it will also be difficultly retrieved. It is in the explicit learning that the learned linguistic knowledge is realized, which, as mentioned above, consists of knowledge as found in dictionaries (Ibid.)".

Finally, procedural learning is also unconscious, it stores data slowly and it is highly related to the development of skills. This learning involves mental representation procedures based on the condition/action principle. They are the ones that, over time, become more automatic. It represents the reliable processes that are called upon in the spontaneity of language production (Ibid.). "These three kinds of learning must work together to enable learners to build up the ability to use a second language for speech production and comprehension" (Ibid.).

Subsequently, the notion of memory is a key term within the psycholinguistics approach to SLA. Towell (Ibid. 51) argues that knowledge must be stored in order to be acquired. Humans would indeed have to store knowledge in memory boxes. There exist three different types of memory boxes, which are the declarative, procedural and working memory (Ibid.). On the one hand we have the procedural memory that handles the procedural, explicit and triggered knowledge in order to expand structured operations in competency-based actions (Ibid.). Furthermore, "it is slow to store but quick to retrieve and will store non-conceptual knowledge, compiling the information it holds into the most 'economic' units" (Ibid.).

The declarative memory "stores triggered, explicit, propositional and conceptual knowledge: it is quick to store but slow to retrieve" (Ibid.). The working memory, on the other hand, comes in between and is central to make the two other function. According to this model, in order to achieve the production or understanding of a language, procedures must be created and, through time and experience, become automated procedures required for real-time communication (Ibid.52).

Another important term in psycholinguistic studies is the notion of automaticity. It represents the procedures that are intrinsically linked to the notion of fluency. However, this is a perceptible manifestation of that phenomenon (Van Moere 2012: 325). In its broadest sense, automaticity indeed "includes psycholinguistic processing speed and the interaction between observed fluency, complexity, and accuracy of the speech act" and is defined as "the near effortless processing of language" (Ibid. 326). Key notions such as formulaic skills, chunking and learner practice define how automaticity develops (Ibid.). This view is based on the assumption that language is a phenomenon that is mainly assimilated through lexicon rather

than syntax and that its acquisition is based on more formulaic skills than on a generative rule-based approach. Ellis (2001 as cited in Ibid. 327) reports on the important place of chunks in the language memory processing. This process represents the acquisition of a series of connections in the long-term memory, which allows the automaticity of speech to expand (Ibid.). As an illustration of this, Ellis claims that native speakers rely on common statements, expressions or even collocations to produce speech. He thus starts from the premise that chunks allow for the automaticity of language production (Ibid.). According to Skehan (1998 as cited in Ibid. 328), language processing and acquisition is carried out via a two-stage process, the "dual-mode processing". This mode consists of alternating rules and memory retrieval. This alternation happens according to the demand of the performance to be produced. Each activity that concerns the use of memory would be related to the spontaneity and automaticity of communicative needs, whereas the rules will be useful in instances of non-urgent communication. Skehan's definition of language performance "reflects the priorities the L2 speaker faces with limited resources, a developing interlanguage, and the pressures of performance goals such as fluency, accuracy and complexity" (Ibid.).

Two opposing perspectives on CAF's relationship to this approach to cognition have been suggested by task-based studies. These two visions contradict each other. It was found that testing these two models was complicated due to non-observable theoretical nature of these variables (Housen, Kuiken & Vedder 2012: 6). On the one hand, the first model is *The Limited Attentional Capacity Model,* Skehan assumes that humans have a limited attention span which implies that they have to focus on a specific aspect of the language that they wish to develop to the detriment of the other two (Ibid.). This model implies that if the learner wishes to focus on complexity, he will do so at the expense of accuracy and fluency. Robinson, on the other hand, assumes with the Multiple Resources Attentional Model that this attention capacity is not limited. He would rather argue that humans have the capacity to focus on a multitude of aspects at the same time and to develop them in synchrony (Ibid.). It is therefore important to note that both models make sense but are difficult to evaluate because cognitive processes cannot be observed. There is no fixed given theory within this approach as it deals with unreachable cognitive mechanisms.

3.3. The limits of CAF

In order to link the psycholinguistic approach to these three variables, they will be firstly listed and defined. Although these three variables have become fundamental in SLA studies, several uncertainties exist for those who undertake to use these variables in their research. First of all, it is important to clarify what is included under the term CAF. It has been shown that these components differ in meaning within the different research studies which has led to inconsistent results. The lack of precision and definitions or the sometimes too selective choice of these variables lead to the difficulty of comparing these studies (Ibid. 4).

Firstly, accuracy is the most translucent variables of the three according to researchers (Ibid. 4; Michel 2017: 54) because it investigates what is commonly called errors, i.e. the gap between the language norm and the learner's performance. However, the identification of these errors is not always simple because it depends on the nature of the errors. It is indeed a challenging task to define what is considered as a language norm. Norms can vary according to social contexts and communities' standards; it is therefore necessary to be able to identify how these varieties of language are considered (Michel 2017: 54).

Furthermore, even if researchers were to agree on what the standards are, it would be crucial to distinguish the extent to which the errors vary from the standards. As an example, we can mention the absence of punctuation, which in some cases does not fundamentally change the meaning of the sentence if forgotten. Such errors should be classified as less serious than, for example, mixing up the word order of a sentence because it alters the meaning. Kuiken and Vedder (2008 as quoted in Ibid.) explained that in order to solve this problem, these authors proposed a taxonomy for classifying errors. There would be three types of errors according to their taxonomy, these three types varying from the least serious to the most serious error in terms of communication efficiency. Foster and Wigglesworth (2016 as cited in Ibid. 9) developed a quantifiable measure for accuracy that would allow the calculation of a score for the degree of acceptability of the total production. Accuracy is therefore also the search for "appropriateness" and "acceptability" (Housen, Kuiken, Vedder 2012: 4).

However, as these authors point out, judging errors objectively is not an easy chore. In performing this task, it is likely that a personal appreciation should have to weigh in the balance. In summary, we can say that holistic scales give an overall idea about the accuracy of L2 productions because general measures allow us to count scores such as the number of errors per clause, but we must also take into account the degree of acceptance of errors and consider a way to delimit them in order to be as accurate as possible in measuring accuracy (Ibid. 10-11). This would imply that weighing accuracy involves some subjectivity.

Towell (2012: 52-53) says that to apply accuracy to the psycholinguistic approach it would require a change in the approach to the norm. SLA approaches are rather relativistic, they identify a norm in relation to native or even to other learners of a higher or lower level norms. This implies that as long as researchers can provide a certain level of systematicity in the learner's interlanguage, they can connect this to standards. According to the psycholinguistic approach, in order for the learner to build up linguistic knowledge, he must be confronted with certain mental representations and therefore have already acquired partial syntactic knowledge in the language. The triggering of these unconscious processes has a real impact only if they are supported by learned linguistic knowledge. In order to assess the learner's level of interlanguage, examples present in the learner's language system will have to be observed in order to identify what types of subsystems he masters. It is therefore normal to observe errors at certain stages of acquisition since interlanguage develops as he comes into contact with the language and mental procedures and representations are created. However, Towell (Ibid. 53) mentions the example of a learner of French who at the beginning stage uses a formula such as "Comment t'appelles-tu?". It is obvious that at this stage of acquisition the learner probably does not master yet the use of pronouns or even the structure of a question and still he uses this formula correctly. It must therefore be assumed that although this formula is used, the learner must be able to use it and unpack the underlying knowledge to consider it is really acquired. Finally, the principle of consistency of accuracy, we can also only consider the learner to be reliable once the necessary procedures have been established in the procedural memory, which will allow him/her to use these procedures in a systematic way in a communication situation.

Secondly, the definition of fluency is multidimensional as well. Ellis & Barkhuizen (2005: 156) discusses two broad categories within fluency studies, namely temporal variables and hesitation phenomena. The notions that define these two terms are the number of pauses, the length of these pauses (temporal variables) and repetitions, false starts, reformulations,

replacements (hesitation phenomena). Hesitation phenomena represent decisions made at the time of performance in a real production context. Housen, Kuiken &Vedder (2012: 5) refers to three dimensions within fluency, namely *speed fluency*, *breakdown fluency* and *repair fluency*. These three distinct notions represent the equivalent of what is understood under the two categories of Ellis & Barkhuizen (2005:156). However, Housen, Kuiken &Vedder (Ibid.) argue that

"fluency is mainly a phonological phenomenon, in contrast to accuracy and complexity, which can manifest themselves (and hence can be investigated) at all major levels of language structure and use (i.e. the phonological, lexical, morphological, syntactic, socio pragmatic level)".

It is important to note, however, that the different fluency measures do not all have the same weight in terms of measuring language proficiency since a learner may have a fairly high rate of speech while taking regular breaks for instance. It is therefore necessary to take a maximum number of variables into account in order to obtain meaningful results. Furthermore, de Jong et al. (2015 as cited in Michel 2017: 56) refers to the importance of recognizing that some of these aspects may be more revealing in terms of personal characteristics rather than as evaluators of L2 knowledge and proficiency. Furthermore, Ellis & Barkhuizen (2005: 156), referring to a study of Wiese, claim that the difference in speech rate between learners and natives is very important regardless of the learner's level of proficiency. Such statements should be taken into account if one wants to provide meaningful results within fluency studies. It is also important to note that fluency generates more relevant results for oral proficiency than for written proficiency. Indeed, it has proven to be more controversial in the analysis of written productions because some measures are intended exclusively for the oral form. They should therefore be replaced by flow and length measurements. It would however be relevant to be able to study the written production process rather than its single finished product in order to identify the mechanisms that the learner uses for its production. There are thus studies nowadays that use keystroke recording software to analyze this process in greater depth (Michel 2017: 56). One could observe that even if accuracy and fluency appear to be less controversial than complexity, there is a number of details that should be taken into account. Within the psycholinguistic approach, the notion of fluency is referred to as

“the outcome of the extent to which appropriate procedures for the processing of the linguistic knowledge (learned and triggered) which has been acquired have been created within procedural memory. Speed fluency will clearly be reliant on procedures for storage and recall;

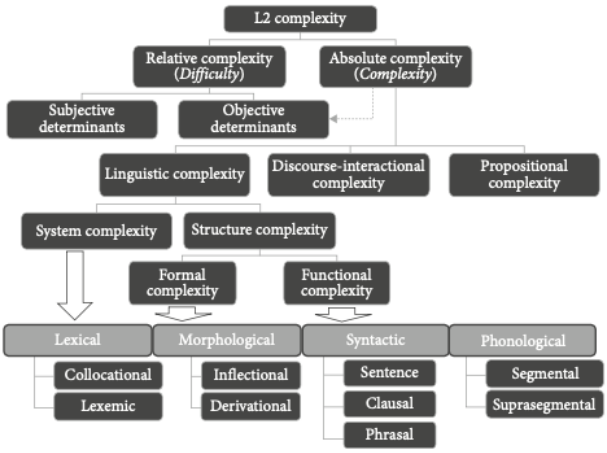
breakdown and repair fluency are related to the extent to which the learner is confident that what has been stored is reliable and the extent to which the learner has also created procedures which can be brought into operation to repair the situation when communication breakdown occurs, for whatever reason" (O'Malley & Chamot 1990 as cited in Towell 2012: 56)

We can therefore consider that procedures will be acquired if fluency is really prevalent in the discourse since it calls for the proper integration of procedures. However, it is important to specify that dealing with the learner's internal processes and procedures cannot be visibly analyzed. It is therefore necessary to examine the capacity for fluency in order to be able to consider whether the procedures seem to be acquired or not mastered.

3.3.1. Complexity

Ultimately, in order to give a full picture of the multidimensional nature of the variable complexity, we refer to the figure 2 below.

Figure 2: A taxonomy of complexity constructs (Bulté & Housen 2012: 23)



Complexity is defined by several meanings and this study has chosen to define them through the model of Bulté & Housen (2012: 23) which divides this term into several sub-categories. The first category, *relative complexity*, concerns mainly the subjective variables that are inherent to learners such as learners' individual differences which will directly influence their learning process as aforementioned in chapter two. These components represent a more sociological approach to the subject. A second dimension of complexity, *absolute complexity*, focuses on more objective data which is not directly related to the learner. It refers to linguistic complexity as such and studies the diversity of the linguistic forms within a learner's L2 performance (Ibid. 24). *Absolute complexity* is also divided into sub-categories, and the one that

is most commonly studied in CAF studies is *linguistic complexity*. The other two subcategories (*discourse interactional complexity* and *propositional complexity)* are very poorly represented in studies so far according to Bulté & Housen (Ibid. 25). Bulté & Housen (2015: 46) argue that linguistic complexity "can be investigated both at the level of the language system as a whole (or of its major subsystems and layers) as well as at the level of the individual linguistic features (forms, items, structures, patterns, rules) that make up such (sub-)systems". These subsystems represent different studies of the language system such as morphology and phonology for instance (Ibid.).

Linguistic complexity is further divided into several sub-categories. They refer to this as *structure complexity* and *system complexity*. These two categories separate *lexical* and *grammatical complexity*. Michel (2017: 53) points out that within lexical complexity itself, one can distinguish many subcategories. We can state for instance one category that calculates, amongst other measures, the size of the lexis. This can also be calculated by using measures such as the type-token ratio. This category represents lexical diversity. Then, we can find lexical sophistication which is the proportion of "advanced" words in a text. Several formulas are used to calculate lexical sophistication, depending on the definition of the label "advanced" (Ibid.52). The infrequent aspect of lexical tokens can be established by using frequency bands such as the 1000 most common words. Finally, lexical density, which is the ratio of the number of lexical words to the total number of word tokens, is used to determine the variety between functions and content words (Ibid.).

The nuances between the vision of *linguistic complexity* as a dynamic component of the learner's system or as a more stable system of the linguistic structures present in his language system is what is at stake here. It is important to state that while researchers identify different types of complexity, it is not always possible to clearly distinguish them. It seems that these constructions are intertwined, it therefore complicates the task of evaluating them. These different points are used to classify and serve as a taxonomy to identify the variables concerned in the study of complexity. This is why some criteria used by researchers are sometimes in contradiction with the classifications made by other scholars. For example, one can cite the difference in classification between characteristics seen on the one hand as complex and on the other hand as simple, in particular with the "s" of the 3rd person singular in English (Bulté & Housen 2012: 26). The nuance between the learner's perception of the difficulty of a given linguistic structure/feature and the level of complexity judged by linguistic studies of these

structures. Michel (2017: 52) indicates that a linguistic structure such as the system of the article in English (the, a/an, zero) could be perceived as complicated to address for the learner whereas linguistically speaking this grammatical system is perceived as relatively easy to understand.

Moreover, it would appear that studies regularly use the same sub-variables when investigating *linguistic* complexity and do not significantly expand their research areas. Robinson and Ellis (2008 as cited in Bulté & Housen 2012: 34) argue that there is a need to focus on specific grammatical phenomena that are often overlooked in order to complement the use of measures of overall complexity. Bulté & Housen (Ibid.) say that "[i]n general, we can say that empirical CAF research has taken a rather narrow, reductionist, perhaps even simplistic view on and approach to what constitutes L2 complexity". Norris & Ortega (2009 as cited in Michel 2017: 53) argue that researchers should avoid calculating “co-linear measures” such as type/token ratio and Guiraud's index which both refer to the same kind of complexity. In addition, if the L2 productions are not of identical length, it is preferable to use measures such as ratios instead of raw frequency (Ibid. 58). However, it would seem that this lack of diversity in research is also due to a lack of adequate tools to measure complexity automatically, since manual data collection requires a considerable amount of time.

The practical analysis will be oriented around linguistic complexity in order to investigate learners' data as objectively as possible. It will therefore be necessary to decide which components we are going to deal with within this linguistic complexity since it is multidimensional as well. Bulté & Roothooft (2020: 2) insist on the fact that measures of complexity are generally taken separately from each other when they are studied. The studies do not emphasize on which measures interact with each other or differ in terms of proficiency levels. The empirical relationships between the selected measures of complexity should therefore be highlighted. Bulté & Housen (2012: 25) argue that most L2 studies do not specify the type of constructions they analyze and do not distinguish them. They even claim that complexity measures are very often calculated in operational terms because they are discussed as quantitative measures. These authors therefore defend three levels that must be made explicit when studying complexity, namely the theoretical part, the observation part and the operational part, in order to avoid reducing complexity to the latter (ratios, etc.). Bulté & Housen (Ibid. 27) suggests that in order to obtain coherent results, “it has to be established first what complexity is (theoretical), how it is or can be manifested in actual language performance (observational), and how these behavioural manifestations can be somehow captured or quantified (operational).

The links between these different levels of construct specification should be as transparent as possible in order for meaningful research interpretations to be made".

3.4. Oral proficiency

Finally, the choice of the competence to analyze in the framework of the practical investigation was strategic. As we have seen throughout this work, the multidimensional nature of SLA and CAF studies has demonstrated the importance of making relevant choices while investigating. However, as aforesaid, oral proficiency is one of the most complex competences to be studied objectively. First of all, Bulté & Roothooft (2020: 13) mentions the cumbersome task of transcribing and analyzing the data manually. Mota (2003: 71) then points out that "comprehension can be more easily assessed than production", which shows the difficulty of the nature of productive skills. It seems, nonetheless, that the main goal of L2 teachers would be to develop this competence so that learners are able to communicate in a real communication context. Mota (Ibid. 70) indeed suggests that oral proficiency seems to be one of the most, if not the most important, of the skills in this view. Studies on the acquisition of oral proficiency, initially studied in first language acquisition research and finally extended to L2 acquisition, are in the majority of cases rare if not incomplete. Dewaele (2000: 17) argues that orality depends on sociosituational and psychological variables and that these factors account for the difficulty of cutting up the spoken chain. Phenomenon such as pauses, elicitation methods or repetitions are common in oral language, which makes it difficult to divide the data into intelligible units. Measures such as type-token ration or Mean Lenght of Utterance (MLU) are techniques regularly used to do so. However, the problem is that a measure such as the type-token ratio is influenced by the length of the data, it implies that the more the sample increases the more the ratio decreases, and this can therefore distort a complexity analysis. Foster, Tonkyn & Wigglesworth (2000: 357) claim that the tendency of researchers in oral language analysis is that they are inclined to omit to make their methods explicit and in case they do, the only examples displayed to readers are simple excerpts in which the problems previously discussed are not found. These authors reviewed 87 studies, including 44 that were lacking a definition of the methods used (Ibid.357). However, when an author speaks of a T-unit as a value equivalent to an utterance, it is not enough to label it but it is necessary to explain and justify what happens to problematic variables such as repetitions for instance.

Foster, Tonkyn & Wigglesworth (Ibid. 362) suggests that T-units are the most common unit used in these studies and it represents a main clause with its dependent clauses (subordinate, associated, etc). However, the type of data with which such measures function is of written nature. It would not be possible to associate a main clause with all its subclauses in spoken language, it is not even always easy to identify a main clause. Foster, Tonkyn & Wigglesworth (Ibid. 360) argues that in order to apply this kind of cut in the oral chain, the T-unit would have to be completely revised. These authors then point out the problem that a great deal of research has been done with the T-unit without ever discussing the problems that could be entailed. Examples of problems in the oral chain are proposed by Foster, Tonkyn & Wigglesworth (Ibid. 363). First of all, the adverbial clause "because". Indeed, it is frequent that this adverb is used at the beginning of a sentence in real-time communication with the function of structuring the speech. It implies that this word is rather used to justify what has just been said, to mean "I said this because" and Foster, Tonkyn & Wigglesworth (Ibid.) specify that "this discourse function is frequently signaled by pause and intonation phenomena".

Another example that can be found in oral data analysis with the T-unit is coordination errors. Some sentences may at first glance constitute the continuity of a main clause. However, thanks to intonation patterns we can notice that some sentences are not coordinated to the main clause. It would appear that theses sentences represent a new beginning and that it should therefore be considered as a new unit. It appears to be common for learners to give up on a subject, creating a type of coordinated sentence that is not originally one.

i.e. "The other woman is very happy now (0.5) and (3.0) just walking away with a gr great smile'." (Ibid. 364)

Bygate (1998 as cited in Ibid.) refers to a common phenomenon in oral language among learners of a foreign language, "independent noun phrase satellite units", see example below.

i.e. Apples are grown in Pakistan These fruits are the same as yours but not gooseberries and strawberries these are not grown in my country (Ibid.).

Indeed, it would seem that the language of these learners is generally a "topic-comment language" (Ibid.). Once again, existing measures cannot provide a justification for this phenomenon. It does not exist in the written language since it represents the flow of speech of the learner who develops and mobilizes his lexical knowledge in order to produce discourse at a given moment.

A final problem noted by Foster, Tonkyn & Wigglesworth (Ibid.) is the phenomenon of interactivity. When a learner learns a new language and interacts, it is generally observed that cooperative work operates between the two speakers. It is possible to observe this phenomenon in the example below. It is therefore not easy to know how to categorize these utterances since they support the end of what the previous learner was trying to utter.

I.e. (Ibid.):

B: I think um (3.0) yeah it's hard maybe he like her but no like a girlfriend, just a

A: Just like a sister.

B: Yeah.

For this reason, Foster, Tonkyn & Wigglesworth proposed a new unit, the Analysis of Speech unit (AS-unit). This AS-unit addresses previously observed problems. This unit will be used in the practical investigation of the data and a full definition and method will be given in order to be as clear as possible with the method of analysis.

Second, the conditions under which oral data is collected is also important. Michel, Kuiken & Vedder (2007: 245) have tested Robinson's *Cognition Hypothesis*, which in task-based studies starts from the assumption that the complexity of certain elements of a task can direct learners' learning towards certain forms of language. On the basis of this theory, Michel, Kuiken & Vedder studied the effects of task complexity and task conditions on the accuracy, fluency and complexity of Dutch language learners. The results indicate that task complexity did lead to more accurate but less fluent production, whereas task conditions, i.e. in the form of dialogue or monologue, revealed that dialogic tasks were more fluent and accurate but structurally less complex. Indeed, in a large number of studies the task that is most frequently used to assess oral competence is the monological performance. Studies suggest that dialogic activities are likely to be the least complex productions of learners and Mota (2003: 73) states that "several researchers have suggested that L2 speech elicited through interactive tasks (...) may be a learner's least fluent variety of interlanguage". Michel, Kuiken & Vedder (2007 as cited in Michel 2017: 62) have indeed demonstrated that "dialogic performance in both populations was characterized by lower grammatical complexity, but higher accuracy and fluency". Gilabert, Barón and Levknia (2011, as cited in Ibid.) argue that dialogic productions are grammatically less complex. Ferrari (2012 as cited in Ibid. 61) also defends the idea that

monological tasks are generally more complex than dialogic tasks on the basis of a three-year study on learners and native speakers of Italian.

However, the *Interactional Hypothesis* (Long 1996 as cited in Drozdzial-Szelest & Pawlak 2014: 92) emphasizes the importance of interaction in the development of oral proficiency. According to this hypothesis, when a learner communicates with a native speaker or with another learner, the interactional process will enable them to point out errors or rephrase each other's words in a more accurate way. This process of interaction allows a kind of negotiation of meaning between the two speakers. Long (1996 as cited in Ibid.) claims that "negotiation for meaning, and especially negotiation work that triggers interactional adjustments by the NS or more competent interlocutor, facilitates acquisition because it connects input, internal learner capacities, particularly selective attention, and output in productive ways".

However, repetition or interactive scaffolding techniques can also be part of the problems encountered when partitioning oral data. Foster, Tonkyn & Wigglesworth (Ibid. 364) claim above all that L2 productions are more complicated to cut than native data. As Bulte & Roothooft (2020: 2) suggests through various studies, research done on L2 proficiency in written productions in learners cannot be applied and generalized to oral proficiency. Vasylets, Gilabert, and Manchón (2019, as cited in Ibid.) studied learner's proficiency performing a similar task but as an oral and a written exercise. The results were that the lexical and syntactic structures were more elaborate in the writing exercises than the oral task. They argue that very few studies have looked at the complexity under different aspects in the analysis of oral productions. The most common measure of complexity in oral research is the morphological complexity.

As we observed in the psycholinguistic approach to CAF, the cognitive processes involved in the realization of spontaneous language such as the role of memory or automaticity are relevant variables but they remain difficult to observe since they deal with cognitive processes. For this reason, it is necessary to rely on more theoretical variables while keeping in mind that the hypotheses will be based on the different studies listed.

CHAPTER 4: Methodology for the practical investigation

4.1. Research project

The development of the communicative approach in language learning studies has enabled researchers to study the acquisition of L2 skills in a framework in which the value of the communicated message is crucial. Along with this new vision in SLA studies, there is the international interest in multilingualism that is constantly growing. It is within this context that the emergence of immersive teaching systems, inspired by the successful Canadian model, has been promoted. Although a system such as CLIL in Europe responds globally to this demand for a communicative approach as well as for the development of language skills within the school's program, it has nevertheless shown both its strengths and weaknesses. It therefore justifies the interest of this master's dissertation in the subject as the present work focuses on the development of the oral competences of learners of English in two distinct systems within the French-speaking Belgian educational landscape: CLIL versus traditional foreign language classes. This study focuses on comparing proficiency measures (accuracy, complexity and fluency) in order to determine the differences between the two groups of learners and to identify potential benefits for the development of their foreign language communicative competence. However, the different measures will have to be calculated quantitatively in order to objectively observe the differences between the two groups. The data that has been analyzed consists of in-pairs dialogues along with their audio files and their transcripts. With the aim of exploring this topic in greater depth, these research questions are the starting point for the further discussion:

1. Are CLIL students more fluent than non-CLIL students?
2. Is the lexical complexity of CLIL learners more developed than that of non-CLIL learners?
3. Are CLIL students' grammatical structures more complex and therefore more advanced in terms of communicative competence than non-CLIL students?
4. Are the CLIL oral productions more prone to error free structures than their non-CLIL counterparts' oral productions?

Different hypotheses are provided to address the above questions. The main assumption will be that, as CLIL provides more input and more opportunities for interactions in the target language, CLIL students will achieve better results in the different areas of proficiency. This

learning environment is conducive to the successful development of cognitive skills and it is likely that it therefore provides better learning outcomes. For this reason, it is assumed that:

1. CLIL students are more proficient in terms of fluency than their traditional counterparts at this level because they are used to use the target language on a daily basis (unlike non-CLIL students).
2. The creativity and therefore the complexity of the structures of CLIL students will be more advanced as they are more cognitively stimulated by an environment that promotes communicative skills development.
3. The structure of the sentences produced by the learners will be shorter than if the data was of monological nature. However, the sentences produced will normally remain filled with complex structures as interaction allows the use of scaffolding techniques to keep up with the conversation.
4. From a lexical point of view, CLIL students will have a higher level of lexical variety than non-CLIL students because they are exposed to more varied and technical vocabulary.
5. CLIL students will not produce error-free language but they will demonstrate the desire to use more complex structures than their non-CLIL peers.

The fourth hypothesis would have required to address the differences between dialogic and monological productions. There were however no monological data recorded to allow this comparison. It therefore means that the hypothesis about the nature of the interactions cannot be confirmed.

This dissertation takes place in the continuation of a large scale-interdisciplinary study carried out in the framework of a concerted research action (ARC project) on the educational, linguistic and cognitive challenges of the CLIL educational system (Bulon et al. 2017 ; De Smet et al. 2018 ; Hiligsmann et al. 2017; Van Mensel et al. 2020).

4.2. Participants and data collection

The corpus is composed of ten in-pair conversations of students in their final year of general secondary school. The first half represents conversations between immersion students and the other half represents conversations between students in the traditional educational

program. The choice to use recorded dialogues between students in the final year gives hope that the communicative competence of these learners is already quite cognitively developed, which will allow for a certain comparison in the data. The corpus was collected as part of the ARC (Action de Recherche Concertée) project. This multidisciplinary project is the result of a collaboration between UCLouvain and UNamur, financed by the FWB. The first part was organized on the basis of written tests in 2015. The data was collected in a context of real communication in classrooms. Some of the tasks were performed on computers in Louvain-la-Neuve. Nearly 500 secondary school students learning English or Dutch in CLIL and non-CLIL settings were followed up in a longitudinal way from 2015 to 2017. The oral data which is used in the framework of this work was collected in 2017 in two schools, one in Marche-en-Famenne and the other one in Wavre.

The corpus contains ten transcribed conversations and three audio files for each group in order to firstly allow transcription and secondly to annotate the corpus with variables such as silent pauses with the help of audio files. The students who participated in the oral tasks had also participated in the written tasks organized in 2015. The various recorded conversations were centered around the theme of an invitation to a party or a trip to a foreign country. They were also provided with five pictures to keep the conversation going. The students had no preparation time and they were recorded using microphones. Data collection was therefore based on a guided elicitation method. As a matter of fact, the collection of corpus data evolved over time and the need for communicative efficiency that we previously discussed had also an impact on this field. Therefore, the collection of authentic data is now favored. However, the need for authenticity is not always easy to fulfill since the simple method of data collection can influence the learner and it can result in unauthentic speech (Gilquin & Granger 2015: 419). Therefore, elicitation methods are employed because they are based on semi-guided techniques that allow a near-natural environment while providing leads to the speakers when they run out of ideas (Ibid.). This technique is based on the learner's response to a pre-established stimulation. This can be done on the basis of a list of words in front of the student or pictures. In this case, photos were provided to the learners and the collection of the data happened in response to a stimulus using a series of five pictures. The task had been clearly stated and the students had to use them while speaking.

4.3. Compilation of the corpora

The spoken data had already been transcribed and controlled by the researchers involved in the project. The next important step involved defining a measure that would partition the oral data in spoken 'units' (a counterpart to the typical 'sentence' in written data). As explained earlier, the AS-unit was selected to accomplish this task because it appears to be the least controversial unit for analyzing oral data as it responds to the complex demands of oral discourse. It is important to give a complete description of how the data was subdivided since this work will follow the guidelines described by the authors of this unit. The different theoretical steps that they established will be paraphrased and quoted in order to be as precise as possible. As the authors state, this unit is mainly syntactic (Foster, Tonkyn and Wigglesworth 2000: 365). Studies on pausing in native speaker data have shown that syntactic units are authentic planning units because a large number of pauses occur at syntactic units (Ibid.). That is why this system was chosen to process the data syntactically. Moreover, it is an analysis that allows to study more than a simple clause in view also of intonation and pause studies that have proven that the speaker regularly plan multi-clause units. The definition of this unit is the following:

> « An AS-unit is a single speaker's utterance consisting of an independent clause, or sub-clausal unit, together with any subordinate clauses associated with either. In the examples that follow below an AS-unit boundary is marked by an upright slash . . . | . . . A clause boundary within an AS-unit is marked by a double colon ::. False starts, functionless repetitions, and self-corrections are put inside brackets {..}. An independent clause will be minimally a clause including a finite verb. » (Ibid.)

It allows sub-clausal units to be counted as units in their own right as they are very often found in oral data. The definition of an independent sub-clausal unit is the following «either one or more phrases which can be elaborated to a full clause by means of recovery of ellipted elements from the context of the discourse or situation » (Ibid. 366).

i.e.:
A: | how long you stay here |
B: | three months. | (Ibid.)

Minor utterances such as yes, thanks, hello, oh, etc., are originally included in the unit because they can be counted as sub-clausal units. However, the analysis is carried out on the basis of dialogic data, so it is advised by the authors to extract such utterances in order not to distort the results of the length of the learners' productions. This work will use the same tagging

system to indicate the different phenomena stated in the definition. Subsequently, the subordinate clauses will be the clauses that consist of at least one *verbal finite or non-finite* accompanied by a complement (subject, object, complement or adverbial) (Ibid.).

i.e. | it is my hope :: to *study* crop protection | (2 clauses, 1 AS-unit)

This measure will be useful since the quantity of learners' subordinate clauses in the corpus will be analyzed. These subordinate clauses can perform different functions. They can either be subject whether they are in the initial or final position in the sentence (i.e. | sometimes it creates problems :: that he knows nothing |). They can also be "verb complementation (object, complement, or catenative verb complementation)" (i.e. | and er they told :: that there there was no food crisis |) or "phrasal post-modifier or complement" (i.e. | still in our country the school and er college students learned the English :: which were er taught to the students before thirty years. |) (Ibid. 367).

Finally, adverbial units that has been previously mentioned with the adverbial clause "because" will be considered as integrated in the unit if they are not problematic. As the authors point out, they are likely to be counted in the initial or in the middle position in the sentence (Ibid.). However, when they occur in the final position in learners' discourse, it is possible that this adverbial clause does not refer directly to what was previously stated, but rather that it is placed as a kind of justification of what is said. One must therefore be careful when dealing with these structures (Ibid.).

Second, there is a problem of classification of the coordinated verb phrases such as "i.e. speaks telephone and look in the other direction" (Ibid. 363). The authors assume that in a large number of cases, even if the subject is omitted, these coordinated verb phrases are new units. However, in order to determine whether they are counted as such, it will be necessary to identify whether the first sentence is not marked by a rising or falling intonation and followed by a pause of at least 5 seconds. If this is the case, these sentences will be considered as independent AS-units.

False starts, repetitions and self-corrections will not be counted in the AS-unit and they will be bracketed to identify them. Repetitions that serve a semantic purpose will be counted as part of the utterance (i.e. | it's a very very bad man |) (Ibid. 368). As far as self-correction is

concerned, it is the final form that is used and counted in the AS-unit. Subsequently, dialogical data can usually contain interventions between the two speakers.

i.e.:

A: | oh that's a big problem
B: oh no!
A: :: because my shop's policy is only :: to give the credits for the return goods | (Ibid.)

In the above example, no AS-unit is granted to speaker B. It will depend on the nature of the interaction and whether it has allowed the other speaker to produce a continuation and whether the production of the second speaker is sufficient to assign an AS-unit to him as well.

Since the corpus had already been checked, the next step involved annotating it so that the data could be coded. Silent pauses were marked with a P and with the value of the pause in brackets to determine whether the pauses were short (1 second), medium (between 1 and 3 seconds) or long (greater than 3 seconds). The duration of these breaks has an impact on different variables, such as silent breaks, for example, which consists of assigning a value between 1 and 3 depending on the length of the break in order to calculate a score for these pauses.

4.4. Choice of study variables

The different study variables were chosen on the basis of theoretical observations discussed in the previous chapters. In order to analyze the proficiency of learners in both groups, a number of variables within accuracy, fluency and complexity have been chosen to best match the requirement of oral data. Whilst numerous potential CAF variables have been identified in the literature (see for instance Bulté & Housen 2012; Michel 2017), we had to limit our choice of variables as this work was carried out in the framework of an MA dissertation. Added to that, it should be noted that this research work has been carried out under exceptional circumstances in view of the worldwide health crisis, the COVID pandemic. As a consequence, the framework of investigation has been limited to certain variables while trying to maintain consistency and address several aspects of CAF.

Accuracy, as previously mentioned, represents the ability to produce language without deviating from the norm and therefore without producing errors. However, as already outlined, identifying errors as fundamentally deviating from the norm is not an easy task and involves a certain degree of subjectivity from an educated researcher in the field. Therefore, in addition to the quantitative data that has been compiled, it will be necessary to use my personal assessment (using my experience as a master student specializing in languages and didactics) to classify the types of errors by addressing those that do or do not affect the quality of communication. In terms of quantitative measures, the average number of total AS-units per speaker out of the total number of error-free AS-units will be calculated. On the basis of examples from the corpus, the way in which errors have been classified will be illustrated.

Three types of analysis were possible regarding fluency. Of the three, repair fluency and breakdown fluency were selected. The first refers to "hesitancy and covers production characteristics related to self-corrections and online monitoring" (Schoonjans 2012: 8). Within these measures, the number of repetitions and the number of reformulations included in the speech can be found. The latter "includes false starts, reformulations and replacements" (Ibid. 22). Breakdown fluency is illustrated by the following definition and this study will use the same way of calculating it.

> “breakdown fluency is operationalized in terms of the total number of filled and silent pauses, each quantified in a separate measure (…). The Number of Filled Pauses (FP) is the sum of the number of lexical fillers (e.g., ‘you know’, ‘like’) and the number of non-lexical fillers (e.g., ‘uh’, ‘hm’). The Number of Silent Pauses (SP) is the total number of silences per transcript, weighted for the length of the silence: short pauses (approx. 1 second) were given a weight of 1, medium length pauses (1-3 seconds) a weight of two and long pauses (over 3 seconds) a weight of 3.”. (Ibid.)

Complexity will be calculated using different measures. To begin with, the importance of lexical richness will be assessed using automatic tools such as Lextutor to measure the proportion of the most common words in learners' speech. It is claimed that CLIL students appear to have a more advanced lexicon than their counterparts in traditional education (see for instance Dalton-Puffer 2008). We will therefore observe whether there is a fundamental difference between the two groups. Another important measure is the use of subordinate sentences in speech. Indeed, subordinates increase the complexity of speech. It would therefore be interesting to analyze the proportion of students that incorporated these structures in their spontaneous oral productions. In order to obtain a quantitative measure, the proportion of

clauses in the number of AS-units will be calculated. It should be pointed out that the subordinate sentences have been distinctly indicated within each AS-unit. Finally, it is likely that we will look at the forms of the different verbs to assess whether or not they have been conjugated and whether or not the tenses used are complex regarding the group. This observation will not be made on the basis of a particular calculation but rather on a global observation of the corpus which will be illustrated with examples.

4.5. Methodology used for the quantitative analyses

The method of analysis of this dissertation is a comparative analysis between two groups of learners, CLIL and non-CLIL students. As explained above, different measures have been selected. In order to allow for comparative analysis, the quantitative data was entered into two different Excel files. The two tables consist of five separate sheets to distinguish between measures of accuracy (one sheet), complexity (one sheet) and fluency (three sheets). The names of the students who participated in the study were anonymized using a coded number. Two different files were created to compute the traditional learner data and the CLIL learner data separately. Each variable, i.e. the number of AS-units, was indicated for each speaker. These tables are presented as follows:

Figure 3: Example of the Excel files

code élève	NB_silentpauses	Nb_repetition	Nb_rephrases	Nb_filledpauses
230501				
230504				
230502				
230510				
230508				
230505				
230515				

The presentation of the results of analysis will be presented as follows in the next chapter. Firstly, the analysis of the fluency data will be presented as described above. Secondly, the presentation of accuracy will be provided and finally, the presentation of complexity data will be presented as well. Once this has been done, we will be able to draw conclusions and answer our preliminary questions and validate or not the hypotheses.

CHAPTER 5: Results and discussions

As aforementioned, qualitative observations are of significant importance for the relevant analysis of the data. As several authors have pointed out (see for instance Housen, Kuiken & Vedder 2012), it is important to consider these variables outside of their quantitative aspect to have a global and relevant overview. For certain variables that will be discussed, case studies will be carried out, focusing particularly on specific productions within the corpus. The observations will therefore be presented before the quantitative analyses are carried out in order to verify whether these two variables correspond to each other.

5.1. Fluency

5.1.1. Silent pauses

In order to calculate fluency values, we were given access to three audio files in each group, three between six students following a CLIL program and three between six students following a traditional program. We were therefore in possession of only six audio files out of the ten initially intended in our corpus. This specific data was therefore annotated with values as silent pauses. In addition to annotating these files, the timing of these silent pauses has also been calculated, which would not have been possible without access to the audio files. However, we could not do the same for the rest of the transcripts.

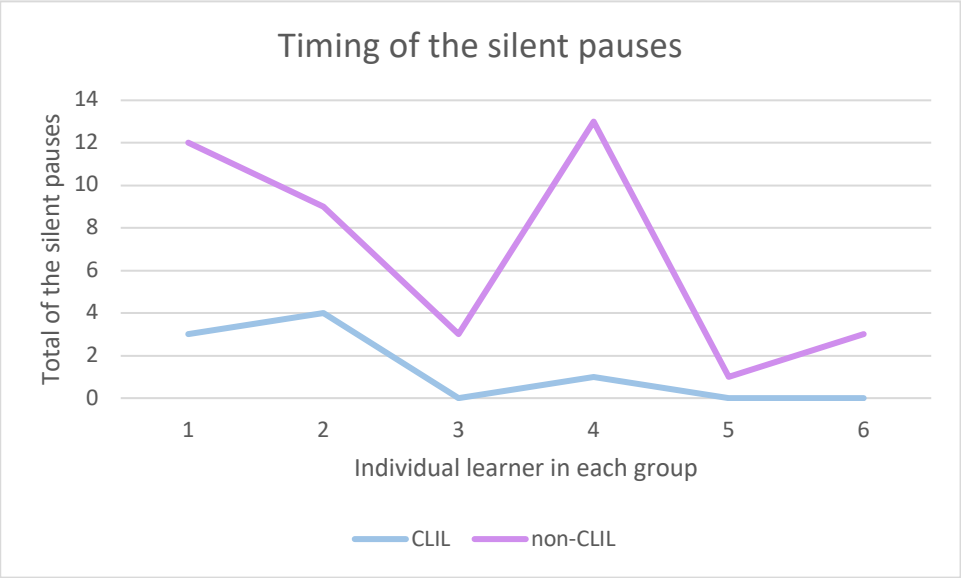
Figure 4: Fluency data of six non-CLIL Learners

code élève	Silent pauses	Number of silent pauses P(>0,3)	Number of silent pauses P(0,2)	Number of silent pauses P(0,1)
230501	9	1	3	0
230504	5	1	1	0
240303	3	1	0	0
240410	12	4	0	0
230508	1	0	0	1
230505	3	1	0	0

Figure 5: Fluency data of six CLIL learners

code élève	Nb_silentpauses	Number of silent pauses P(>0,3)	Number of silent pauses P(0,2)	Number of silent pauses P(0,1)
230313	3	1	0	0
230311	4	1	0	1
230312	0	0	0	0
230316	1	0	0	1
230408	0	0	0	0
230402	0	0	0	0

Figure 6: Timing of the silent pauses in both groups



As can be seen from the graph on the previous page that represents the proportion of silent pauses, a clear difference can be noticed between the two groups. By adding up all the breaks of the students in each group, the average was calculated. It represents an average of 1.3 seconds in the CLIL group while the average of traditional students 5.5 seconds. This average was calculated by adding up all the different pauses (short, long and medium). It is important to look at what this average includes by looking at Figures 4 and 5 above to see the different types of breaks represented. The difference is particularly noticeable in one case with a total of 12 representing 4 pauses of 3 seconds or more. In this case, it means that the pauses are relevant markers of dysfluency since they involve long breaks. It can be noted that in the CLIL group, the maximum is 4 in which case the learner marks only one pause of 3 seconds or more. The pauses of two seconds or less are globally considered as being less significant because they are considered short and therefore less relevant in terms of dysfluency markers (Schoonjans 2012 :22). It can therefore be noticed that the timing of the silent breaks is not particularly significant in both groups. However, this does not imply that the speech of a student with a high score in silent pauses is totally saccadic and that these pauses hinder comprehension since it is possible that the student regularly marks short pauses. However, in the case where long pauses are frequent it will be more quickly noticed that the speaker is less fluent and that he therefore appears as less proficient. As a matter of fact, in non-CLIL students 230501 and 230504' speeches, it can be observed that although the silent pauses are more frequent than in CLIL students' speeches, comprehension is not necessarily heavily impaired. These pauses might simply

involve the process of searching for ideas while keeping up with the conversation. Indeed, these students did not have preparation time.

In contrast to these two students, the overall impression with non-CLIL students 240303 and 240410 is that the level of fluency is impacted by the lack of knowledge and/or understanding. In this conversation, one of the two speakers regularly takes long breaks. The average for the second speaker is not very high. Listening to this discussion requires concentration to follow the course of what is said because they regularly pause or the meaning of part of the sentences is not always clear. The researcher (<R>) who was present during the recording of the data intervenes at times to either restart the discussion or ask for some clarifications so that the conversation can keep up. In the course of the conversation between students 240303 and 240419, she intervened to put the conversation back on track and give them new ideas. However, she had to repeat her question so that the students could answer correctly to her question. Below we can find the example of that specific moment:

<R> who invited you to the party ? </R>
240410: mm in first. eh her party and after my party
<R> yes I know but who invited you ? is it someone from school is it a friend ? </R>
240303: ah yes mm and who invited you ?
240410: ah mm my best friend

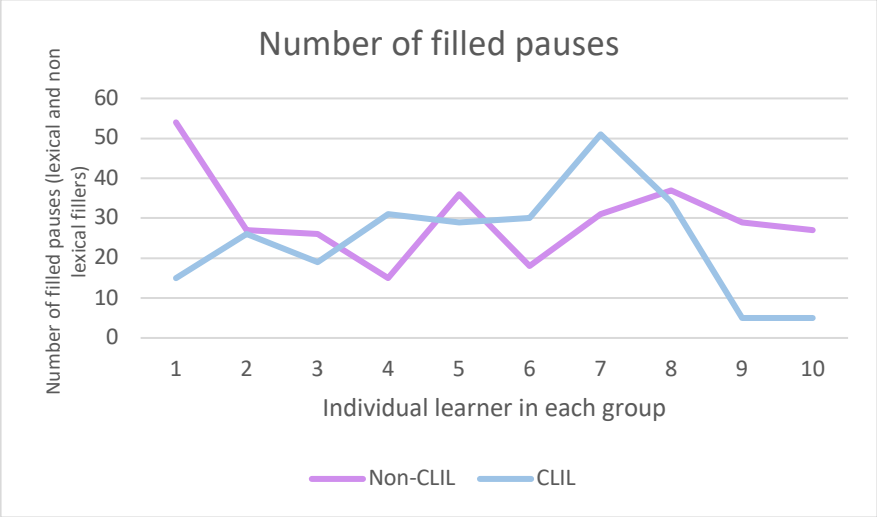
It can be seen in the above example that the right information was not understood by the two learners. After the repetition, one of the two students understood what the researcher had asked them, which was information about who invited them to the party. The overall impression is that the silent pauses have an influence on the impression of proficiency in the speech of these learners, but it is not indicative of their actual knowledge or understanding. It is clear that in the case of the two traditional students whose average number of silent pauses is significantly higher that the impression of competence appears to be lower. This is due to the many times that these two students have to search for their words. However, although the average of the two groups of learners is significantly different, it is difficult to draw conclusions. In particular due to the small size of the sample studied, it cannot be representative of all CLIL and non-CLIL students throughout the French-speaking part of Belgium. Nevertheless, what the data allowed one to see is that

there is certainly a difference between the two groups in terms of fluency. To draw more conclusions about the fluency of the two groups, we will look at other variables that can be applied to the group as a whole.

5.1.2. Filled pauses

The number of filled pauses is a number that was calculated without the help of the audio files because the corpus had been carefully annotated beforehand to successfully carry out this operation. The variable was therefore calculated for the whole corpus. The number of filled pauses was counted by summing the lexical fillers and the non-lexical fillers present in each learner's speech. A graph representing the two groups is shown below.

Figure 7: Number of filled pauses in the two groups

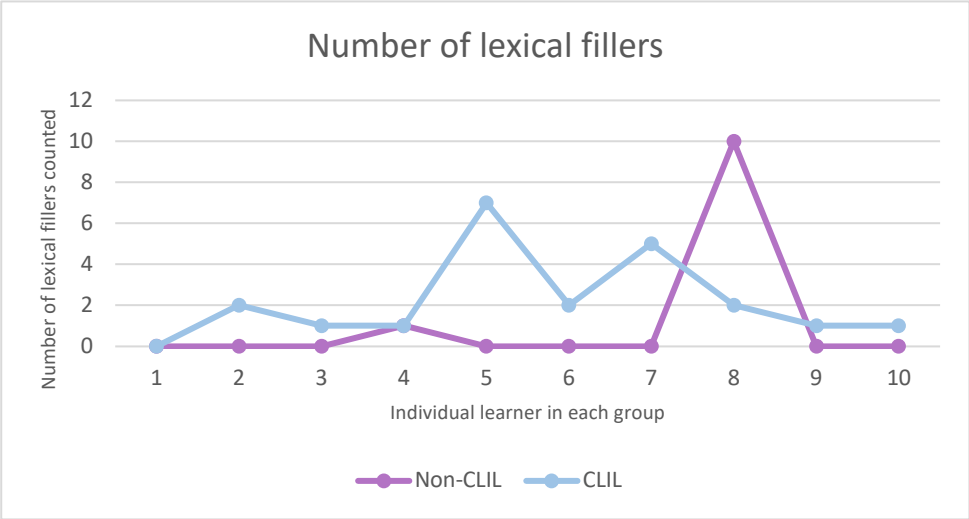


From the graph, it can be seen that the highest number of filled pauses was produced by the non-CLIL group and the lowest number was produced by the CLIL group. However, the curve is variable since it can be noted that in the CLIL group there is a peak in the highest numbers as well. Furthermore, it should be relevant to look more specifically at the repartition of lexical and non-lexical fillers present in both groups.

In order to assess whether the difference between these two groups was statistically significant, the p- and t-values and the mean number of the filled pauses was calculated. The average for the non-CLIL group is 30, while the average for the CLIL group is 24.5, thus slightly lower than that of the first group.

At the level of the p-value and the t-value, this calculation was carried out automatically using the Social Science Statistics website. The results were presented as follows: The t-value is -0.98004. The p-value is .170027. The result is not significant at $p < .05$. It is therefore observed that although the mean of the two groups is different, this difference is not significant. We cannot therefore say that the use of filled pauses is not fundamentally different between the two groups.

Figure 8: Number of lexical fillers in both groups



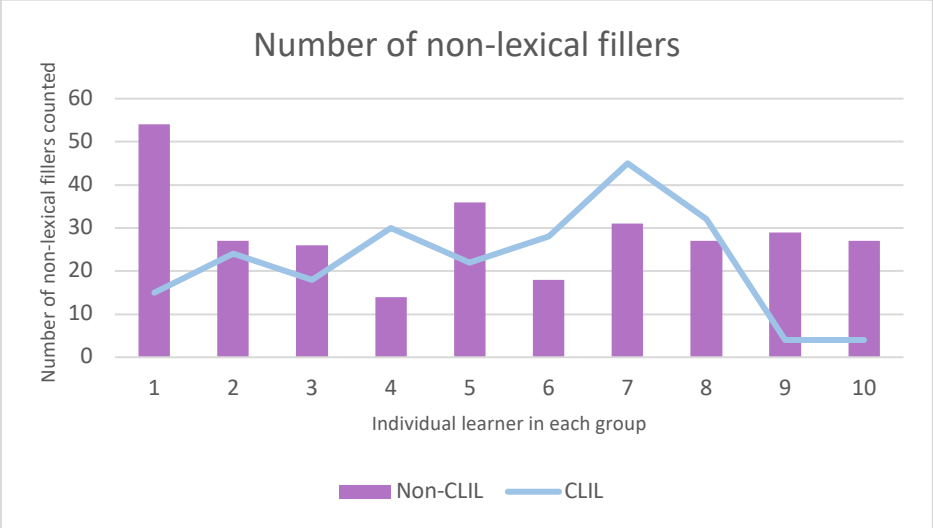
The common forms of lexical fillers found in the corpus are "you know, I mean, like, etc.". These forms have been in general little used in both groups. However, student number 8 in traditional education from the graph has a high ratio since he produced ten lexical fillers in his speech. These speech phenomena perform different functions.

"Lexical fillers are special in the sense that they often fulfill more than one function at the same time. In addition to 'playing for time' they can fulfill social, interactional discourse, and symbolic functions, such as engaging the addressee, yielding the floor, asking for feedback, stressing the content of an utterance, making it sound more friendly" (Rieger 2001: 81).

According to the study of Rieger (Ibid.), such phenomena - as well as non-lexical fillers - are only scarcely encountered in the beginners' speech who tend to merely produce silent pauses. According to the same author, studies on native speakers' skills have revealed that they "use a variety of fillers to fill their hesitation pauses, such as the lengthening of sounds, quasi-lexical fillers (uh, uhm), lexical fillers (well, you know, etc.) and repetitions" (Ibid.). As far as this study is concerned, the phenomena of hesitations such as " uh, eh,

etc." are called non-lexical fillers, not quasi-lexical fillers. The following graph represents the quantity of these non-lexical fillers observed in the corpus.

Figure 9: Number of non-lexical fillers in both groups



Most of the lexical fillers observed in this corpus are "eh,mm". It can be observed that these phenomena are more common in the traditional education students' corpus. However, it appears that they are indicative of dysfluency. Rieger (Ibid. 84) concludes by saying that lexical fillers are used to make up for the over-recurrent use of non-lexical fillers. Indeed, "quasi-lexical fillers are often thought of as negative and undesirable flaws. Too many quasi-lexical fillers in one's speech apparently convey the image of an ill-educated, disorganized person. Conversely, idiosyncratic fillers like all lexical fillers reflect favorably on the speaker since their additional social, discourse, and interactional functions make a person's speech more friendly and engaging" (Ibid.).

On the basis of various graphs, we can say that overall the use of lexical fillers is more frequent among CLIL students, although the use is not necessarily high. The traditional students do not seem to employ them, except in one specific case which goes far beyond the use of lexical fillers by CLIL students. However, the use of non-lexical fillers seems to be much more frequent among traditional students than among the CLIL group. As Rieger (Ibid.) reports, these non-lexical fillers give an unfavorable impression on the learner's speech and are considered as signs of non-proficiency. In order to have a critical opinion on this matter in this dissertation, we will study a few cases in depth so that we can draw a more informed conclusion.

5.1.3. Case-study

In order to compare the two groups, a few students from both groups were selected based on the number of filled pauses found in their dialogues. The students with the most filled pauses in each group were selected so that the two groups could be compared with their extremes. In decreasing order, CLIL and non-CLIL students were ranked as follows:

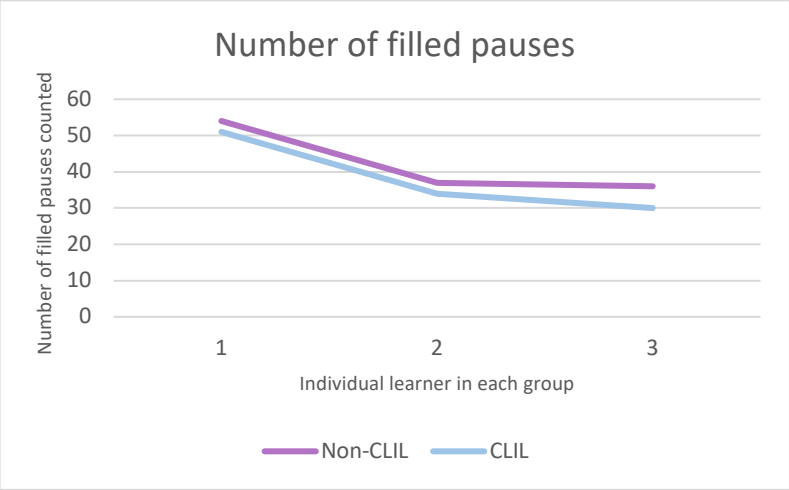
Table 1: Highest number of filled pauses in the CLIL group

CLIL students	Number of filled pauses
230414	51
230411	34
230316	31

Table 2: Highest number of filled pauses in the non-CLIL group

non-CLIL students	Number of filled pauses
230501	54
230503	37
230508	36

Figure 10: Number of filled pauses of three members of each group.



The preceding graph represents the pattern of the two groups. It can be noticed that they seem to have more or less the same numbers in their highest number of filled pauses. However, in view of the more limited access to the audio files, we did not have access to all the files of the selected dialogues, it was difficult to give a general impression on the level of fluency of the students and to know whether or not these filled pauses could interfere with the students'

competence when we could not hear what they were actually producing. However, from the transcripts, we found that the CLIL students' filled pauses seemed to interfere less noticeably with fluency in general. Indeed, as an example, we can state that many of the occurrences in the speech of 230414 are "yeah", which does not give the impression of a significant break but rather an impression of approval.

i.e.:

yeah camping is a bit rude and . I don't really like camping eh on a eh on a . other picture eh there is plane so . I guess we'd go by plane {to}: to Thailand that's obvious I think:: {we can't} eh we can't go {by} eh by car or things like that . eh

Whereas in the speech of student 230501, the pauses seem to be much more disturbing for the fluidity of delivery of speech.

i.e.:

|yes eh but I have eh also a party |.| and eh {I } I wanted eh ::that you come with me too |. mm <laugh> eh it eh

However, it is important to point out that these are only observations made on the basis of transcripts and not based on direct listening. Moreover, this sample used is quite small. It is therefore not valid to take this observation into account and consider it as a given and determining factor for the fluency of the two groups. Based on the few audio files we had access to, differences have been certainly noticed between student 230316 and student 230501 because the first counts 31 filled pauses and the second counts 54 filled pauses. As the two are not equivalent in terms of their number of filled pauses, it is therefore not correct to judge the overall fluency skills of the students on the basis of two audio files.

5.1.4. Conclusion

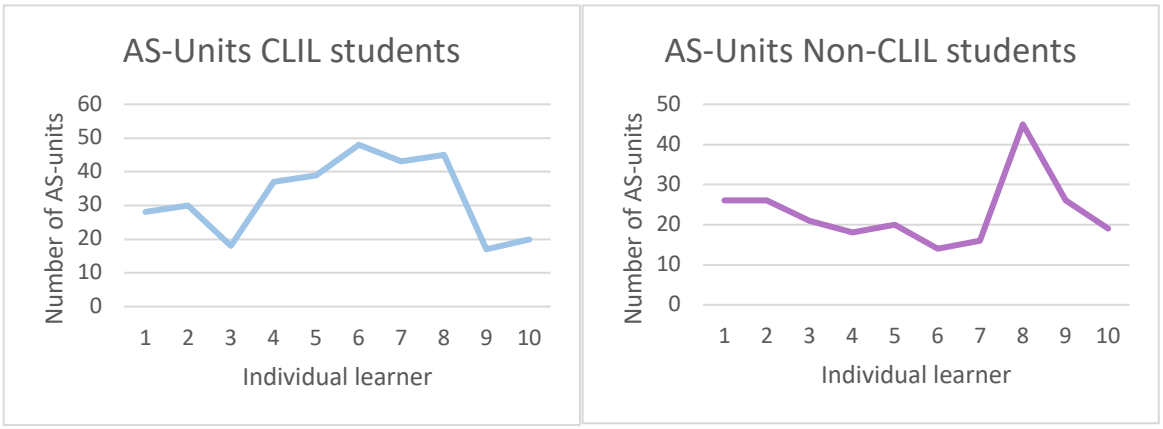
On the basis of the findings, it can be concluded that overall CLIL students tend to be slightly more fluent than non-CLIL students, particularly because the number of filled breaks is slightly lower. In order to answer the initial hypothesis to determine whether CLIL students are more fluent than their counterparts in traditional educational settings, the discussion will be moderated. Generally speaking, in view of measures such as silent pauses and filled pauses, the CLIL students seem to slightly exceed the abilities of the second group. As specified, in terms of filled pauses although a slight difference was noted between the two groups, it was determined that this difference was not significant. It is therefore not correct to draw conclusions and say that this difference exists. However, it is important to specify that the

access to the audio files, as previously explained, has been restricted due to the current health crisis. Consequently, it was not possible to listen to all the conversations and to judge subjectively whether the data corresponded to an overall impression. For this reason, the hypothesis is partially validated, which means that according to this corpus, CLIL students are partially more proficient in terms of fluency than their traditional counterparts at this learning level.

5.2. Accuracy

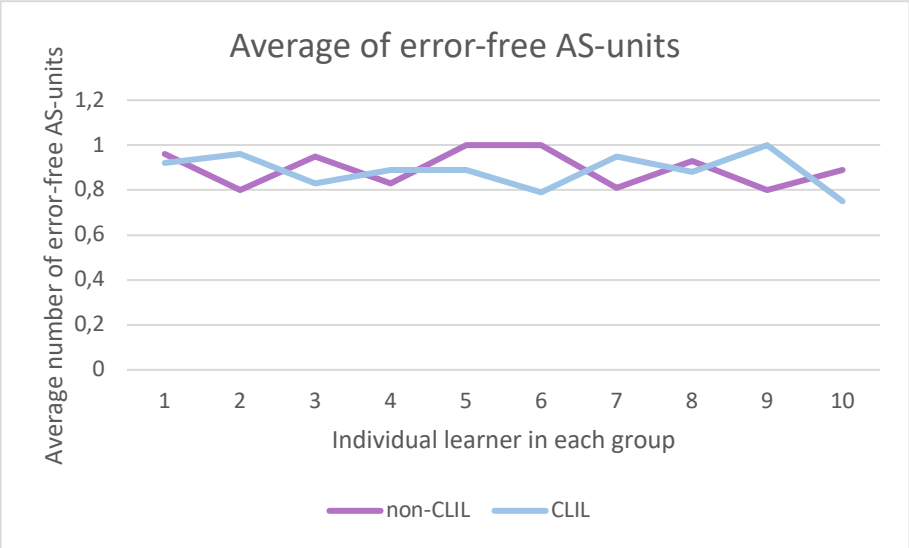
Accuracy, as aforementioned, is the study of deviations from language standards. However, the difficulty of assessing the deviation forms from a given standard is represented by the obligation of researchers to consider errors as being more or less serious in terms of communicative efficiency. In order to study the data, the number of AS-units in each speaker's speech and the number of error-free AS-units were calculated in order to obtain an average of the number of learners' errors. Another interesting value was the average number of words per AS-units in order to see the potential length of the speech and thus the ability to assess the greater or lesser possibility of learners to make errors. After presenting the statistical data, we will utter a personal judgment based on these errors and a few examples from the corpus.

Figure 11: AS-units in both groups



As one can see from the two graphs, it seems that the production of AS-units is generally higher in the CLIL group. It implies that they produced more independent sentences while speaking than the second group. However, it should be noted that these sentences can also represent sub-clausal units, which are quite frequent in speech and which were determined as part of the AS-unit.

Figure 12: Average of error-free AS-units in both groups



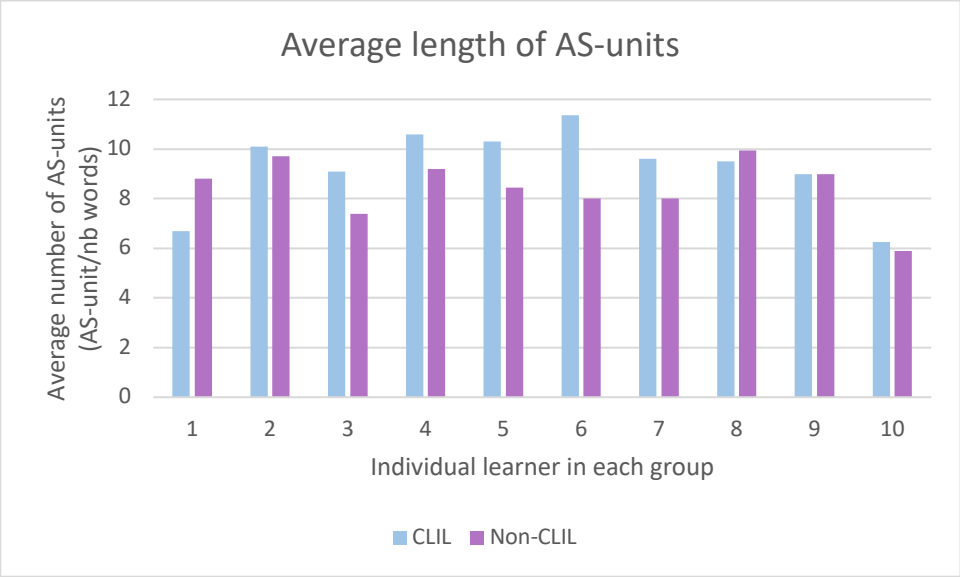
In this graph, it can be noticed that the average of the error-free AS-units of both groups. It appears that non-CLIL students make fewer errors than their CLIL colleagues. This could be explained either by their ability to make fewer errors in general or by the length of their productions. If non-CLIL students tend to produce shorter AS-units than their CLIL counterparts, it might make sense that the latter commit more errors.

For instance, the ratio of student 240410 is better than that of his classmate 240303. However, as one can see from an external point of view, student 240410 produced shorter sentences and did not experiment with many complex structures.

240410: mm ... I'm invited eh .. in a party . {with} eh ss with a friend mm .. mm .. |{they} they are eh .. <laugh> music and fireworks {they} they are eh .. <laugh> music and fireworks {cham} champagne oh oh no I prefer my party because eh {the/} there is a <foreign> limousine </foreign

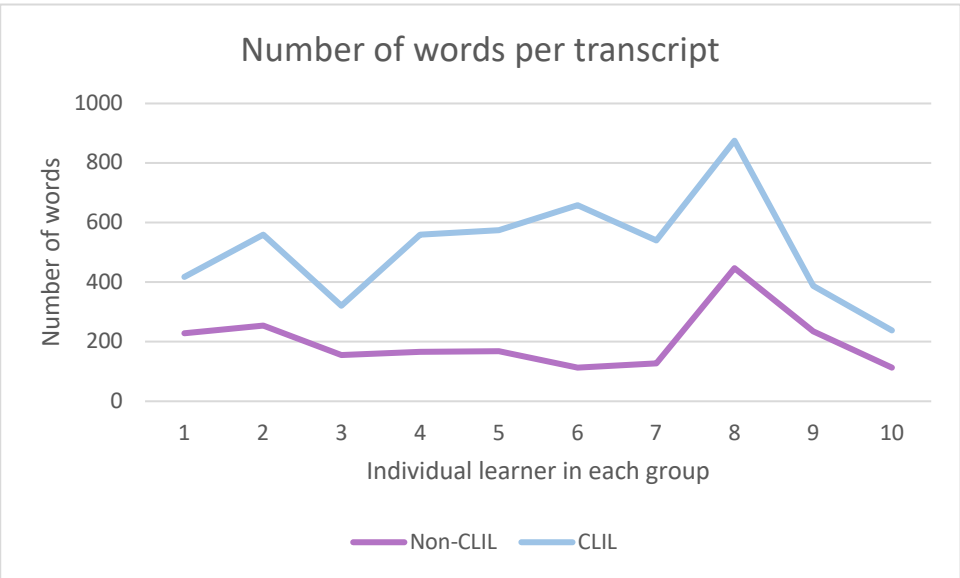
240303 but if you want {after} after {the: }. the movie and eh monopoly party :: we will go eh {in}. eh in a discotheque to dance |

Figure 13: Average length of AS-units in both groups



As we can observe, the length of AS-units of CLIL students is slightly bigger than the one of non-CLIL students. It can be observed that in two cases the traditional students exceed the production of the CLIL group. However, it does not appear to be a striking difference in most cases. We cannot jump to conclusions and we cannot claim that CLIL students make more mistakes only because they tend to produce longer units based on these findings.

Figure 14: Number of words per transcript in both groups



However, this graph shows that in terms of words, CLIL students distinctly outperform traditional students. It is possible that this phenomenon is due to a higher fluency rate. It can be observed that although in terms of average length of AS-units, the difference between the two groups is not significant, the overall productions of the CLIL group are longer than the productions of the non-CLIL group. It also depends, as it will be discussed in the next point, on the complexity of the sentences constructed by the learners. CLIL students could have a fairly high rate of speed while they are producing simple sentences and thus produce a larger number of AS-units.

Returning to the accuracy of these two groups, it can be observed on the basis of the data that although the traditional students seem to have a lower average of errors in their conversations, it is still necessary to take a look at the types of errors of these learners in order to judge whether the quantitative results are representative or not of the reality.

5.2.1. Case study

In order to illustrate the mistakes that have been made throughout the corpus, several varieties of mistakes were selected. As a starting point before classifying certain types of errors, two students are compared. The first is the student 230408 who produced 48 AS-units and 38 of them without mistakes. Among the non-CLIL students, the student 230503 who produced 43 AS-units without errors was selected.

Table 3: Examples of errors from student CLIL 230408

1. I won {a:} a vacation for going to Italy
2. and there is women with eh cock
3. In Thailand there is eh change women
4. yeah I'm agree on that
5. and eh we have a tent to make camping eh with a fire during the night and all that
6. it is eh {more} more huge than your swimming pool
7. in this lake there is also fishes
8. but you know Thailand is {very} very polluted so thanks to the <X> many polician
9. {and} eh and if it is a good lake:: you don't have bacteria on in it
10. so maybe {we} you will {ta/} / find eh a piece of the airplane that disappear

Table 4: Examples of errors from student non-CLIL 230503

1. or it’s just the food that you > for tourists ? 2. but me it’s Italy . so there is sun too 3. it’s would be good a mix of the two

The first ten examples are those of student CLIL 230408. We can notice at first sight that the learner's mistakes are for example the omission of the plural form in sentences 4, 7, and 8. This type of error, which is certainly a serious basic mistake, does not globally impede comprehension and therefore communication. In sentence 4, it is interesting to observe that the learner tried to use a fixed language formula such as "to agree with somebody" but when he used another preposition, he made a mistake. Even though this is to be considered as serious, one can notice an effort to use formulaic language. If the learner had not changed this preposition, this phrasing could have been an indicator of the quality of his production. For the second learner, overall comprehension in sentences containing errors is more complicated. Example number 2 is quite similar to a basic structure borrowed from French. In the third sentence, an effort to use a modal verb can be noticed. However, the addition of the "is" in front of it shows that the learner does not know how to use this form correctly. Indeed, as it will be pointed out later, most non-CLIL learners do not use modal verbs. In this case it can be noticed that the effort has been made but that the final form is wrongly used.

This brief comparison, before going into the more specific errors in the corpus, shows that it is interesting to look at the overall production to judge on how serious errors should be considered and whether or not they strongly impede comprehension. It can be seen that even though the CLIL learner produced more errors in the above case, it is easier to follow what he intended to mean than his non-CLIL counterpart.

5.2.1.1. Pronunciation mistakes

First of all, there are several pronunciation errors in the productions. In some cases, the learner did not pronounce the tonic accent at the appropriate place in the word, resulting in a pronunciation that closely resembles that of the mother tongue. In the following example, the French word and the English word are non-homophone homographs. It means that what differentiates the two words is the pronunciation. The meaning behind the word is the same but

the pronunciation is different. For this reason, when annotating the corpus, the researchers identified this word as belonging to a language other than English since it was mispronounced by the learner. It should be noted that since the learner is conversing with another Francophone learner, this will not prevent the understanding of the word "champagne" or the word "horrible" since they both speak the same mother tongue. This mistake could lead to a higher cost if the learner was conversing with a native speaker.

Student 240410 (Non-CLIL)	I like pizza but I prefer <foreign> champagne </foreign> It is horrible

Other pronunciation mistakes have also been found but in this case, they alter the original meaning.

Student 240303 (Non-CLIL)	no we will talk about . boys: and eh and girl :: that eh we ate
Student 230510 (Non-CLIL)	I like city trips that hallow me to see a lot of eh sceneries and eh monuments

5.2.1.2. Grammar mistakes

In terms of sentence structures, a common mistake that occurred in both groups is that they tend to not finish some sentences that they started at some point. The sentence remains unfinished and the meaning never reaches the second speaker. Most of the time, it happens when the sentence is in final position and when the second speaker takes his speaking turn.

Student 230501 (Non-CLIL)	yes eh but I have eh also a party |.| and eh {I } I wanted eh ::that you come with |. mm <laugh> eh it eh …
Student 230503 (Non-CLIL)	because I'll be mad at the end of the…

Student 240303 (Non-CLIL)	you don't sing :: cause {you}: you don't have eh a[ei] good voice | or <laugh> but ….
Student 230311 (CLIL)	eh there is a barbecue .. and eh …

As it can be observed in these above examples, the beginning of a sentence is initiated but it does not reach its end. This characteristic found in the corpus is probably simply due to the fact that in oral language, sentences are not cut out as radically as in written data. It is therefore certainly interesting to note that this is one of the characteristics that has been observed, but it is not directly related to the belonging to a specific group of learners. This kind of phenomenon is also probably due to the fact that the speaker, while thinking, has started a sentence and has been interrupted by the second speaker since the latter was ready to interact.

5.2.1.3. Words omitted

Another type of error is the omission of elements in a sentence such as the subject, the verb or one of the complements. As Foster, Tonkyn & Wigglesworth (2000: 364) pointed out, it is common to find this type of phenomenon in oral speech. However, as much as this factor could be problematic for the cutting of data it does not impede global comprehension of the message. Student 240117 omitted to use a subject, but it is easy to understand what he meant. This kind of phenomenon is typical of oral communication. It should also be considered that in oral production external factors are constantly intervening, contrary to written production. In this corpus, we can mention, for example, occasions when students that have run out of ideas use the provided five pictures. It can happen that the conversation goes from one end to the other without any clear link between the two. This is one of the characteristics of oral productions.

Student 240117 (CLIL)	yeah ok should be good

5.2.1.4. Use of the mother tongue

It can also be noted that learners sometimes switch back to French when they are unable to communicate what they wanted to say. Referring to the corpus, the group that seems most likely to use their mother tongue is the non-CLIL group, with a proportion of three occurrences of French out of five conversations. In the CLIL group, we find two occurrences out of five. In most cases, these learners switch back to French to communicate more easily and waste as little time as possible on the search for a word. In all cases here, they get help either from the other speaker or from the researcher.

> 230501: and eh if you come at my party you you: haven't haven't to .. to . je sais plus comment on dit mm to bring money because eh . everything is eh is eh …
>
> 230504: free ?

> 240410: no eh I. <sigh> . I . <foreign> { pfou je sais pas comment on dit ça . mais je l'ai déjà vu }

> 230408 : we wouldn't be so: . eh I'd say eh . eh je ne sais même pas le dire en français
> 230402 : homesick . homesick?

In addition, sentences starting with the structure "me ..." are frequently used. Regarding this type of error, the transfer process that we discussed in the previous chapters of the theory can be mentioned. Since the mother tongue of these learners is French, it was more than likely that transpositions of French language structures in the L2 were to be found in this corpus. In terms of comprehension, such mistakes do not prevent communication and the learner's message can still be conveyed. However, this type of error can be significant of the lack of knowledge of the real structure and use of English since such formulas will never be found in native productions.

> Student 230502: yes . me I prefer: mm going to eh . eh . city trip

5.2.2. Conclusion

In conclusion, it can be assumed from the analysis of the accuracy data that it would seem that the interlanguage skills of CLIL students are indeed not error-free, but that these students demonstrate the desire to use more complex structures than students in traditional courses. These data confirm the initial hypothesis. However, it should be kept in mind that these students, being hardly or not at all confronted with grammar courses in the target language, produce errors that may seem serious at an advanced level. However, it has been noted that some of these errors do not prevent the understanding of the initial idea.

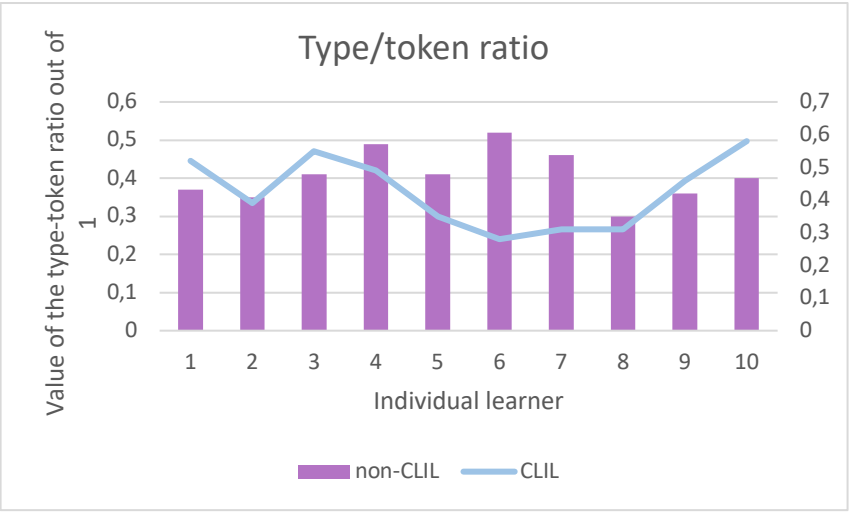
5.3. Complexity

Finally, the last variable that is investigated is complexity. As defended in the theoretical part of this work, this variable is the most controversial variable in the CAF triad. However, it has been decided to study some measures within complexity, but it will not be possible to get a complete view of it because of the time-consuming nature of this analysis.

5.3.1. Lexical diversity

First of all, in order to analyze complexity, we will look at lexical diversity. Lextutor is a website that allowed us to extract the type-token ratio of each learner automatically. The type-token ratio represents the number of different words in the selected text divided by the number of words in the same text.

Figure 15: Type/token ratio of both groups.

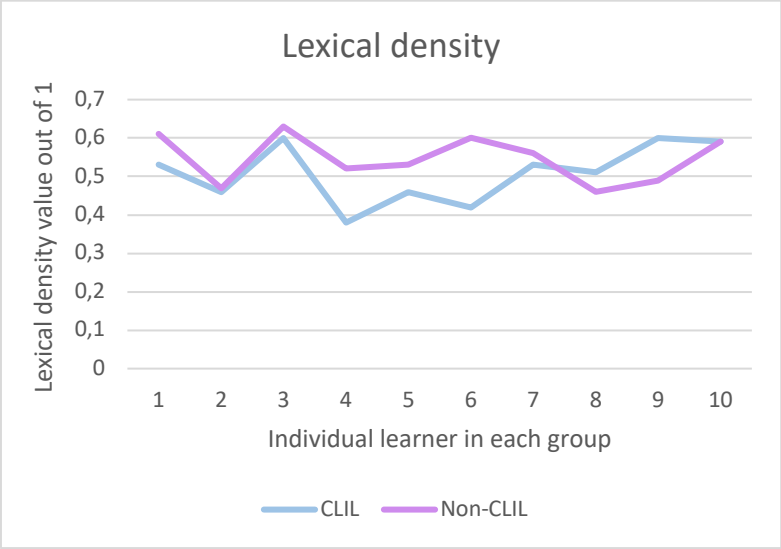


This graph allows us to see that the type-token ratio is globally similar in the two groups. Even if at some places the CLIL students outperform the non-CLIL students, the inverse phenomenon also occurs at other places in the graph. However, as noted earlier, measures such as the type/token ratio are influenced by the length of the productions. This could disadvantage the CLIL students who tend to have longer productions than the other group in general.

5.3.2. Lexical density

Subsequently, the lexical density was also calculated automatically by the website Lextutor. This value represents the number of content words out of the total number of words. It should be noted, however, that most of the conversations were strongly influenced by the images in front of the students. In fact, vocabulary terms are used quite frequently, and it makes sense since we can talk about topic vocabulary words. They are indeed influenced on the one hand by the topic imposed and on the other hand by the pictures in front of the learners.

Figure 16: Lexical density of both groups



This graph shows that in terms of lexical density and thus in terms of the density of content words over the total number of words, non-CLIL students exceed their CLIL colleagues. It could imply that non-CLIL students tend to use a more developed vocabulary than CLIL students. However, this difference does not seem to be significant.

5.3.3. Vocabulary

With the help of Lextutor, we were also able to examine the proportion of words used by learners in corpora of the most commonly used English words. It was noted that most of the words used in the texts were included in the 2000 most commonly used words in the English language. In this corpus, it was noted that there was no fundamental differences between the two groups in terms of vocabulary acquisition. However, it should be noted again that vocabulary is influenced by the topic of the discussion and that unless one investigates academic data, one rarely finds words that exceed the 2000 most used words in a corpus.

5.3.4. Grammatical complexity

5.3.4.1. Tenses

In terms of the tenses used by the two groups, it was observed that the ones present in the speech of CLIL students are more complex. Indeed (see examples below), we found in six out of ten CLIL students' speeches forms of the perfect present, whereas in traditional students' speeches the most elaborate tenses include the continuous present, the past, the present and the future. These tenses were also used in CLIL students' interactions. No occurrence of present perfect was found in non-CLIL students' speech. For instance, in the dialogue between the two non-CLIL students 240303 and 240410, while the researcher formulated a question in the present perfect to restart the conversation, one of the two students tries to repeat the question but neglects to put the past participle in the correct form. There is therefore no correct use of this time in non-CLIL speeches.

Student 2303101 (CLIL)	{fine} fine I've been invited {to a : to a : .. } to a party tonight
Student 230411 (CLIL)	but you said :: that you've been eh ah no you didn't {sorry sorry not} sorry you've been to Italy not to Thailand

Student 240303 (non-CLIL)	you} you have already see it or not ?
Student 230408 (CLIL)	the trip I've won
Student 230313 (CLIL)	I've also been invited to a party
Student 230414 (CLIL)	I've eh won a kind of lottery eh

Although the structures of the present perfect are all used properly by CLIL students, this does not mean that they always use all tenses correctly. Indeed, it should be remembered that CLIL students' learning is particularly centered around oral communication, which could explain the difficulty CLIL students experience with grammar and therefore the difficulty with producing error-free language. Indeed, they assimilate structures in their interlanguage in general without knowing exactly how to use them properly.

5.3.4.2. Modal verbs

The use of modal verbs such as "could, would, might" have been observed in the corpus. We therefore looked at the proportion of use in the two groups. It seems that modal verbs are found in six out of ten speeches among CLIL students (see list below). Whereas in the ten non-CLIL dialogues, we found only three learners who used them.

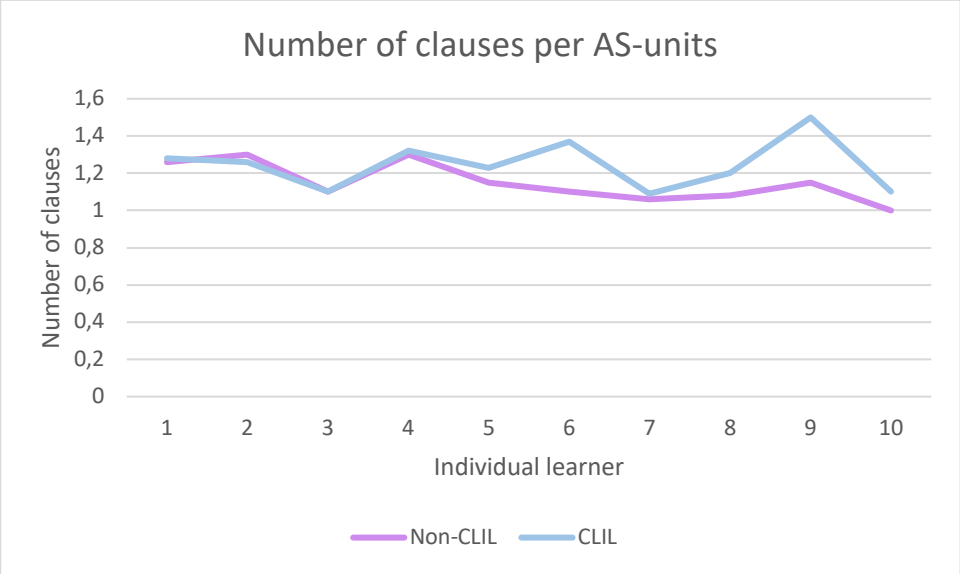
i.e.:

CLIL students	
230313	I think we could go to yours because eh I really love listening to DJ's music and also swimming so I think it could be great
230316	but we . { might . } eh . might not have the[i:]. possibility ::to get to know . people :: like we could in yours :: because it's eh biggerand . yours seems smaller so we could . talk more in yours so: I think:: it might be cool
230311	so we could go eh swim eh there is a barbecue .. and eh so: you might maybe prefer
240117	I would maybe enjoy it
230411	hot weather I would say |</ how would I find this American ?
230414	so we would go skiing ? and {would you what} . where would we:: I don't know sleep ? for example
240408	so I think it would be great and . should be good with eh the adventure trip yeah but . you know if you don't like sports you should make sport to get fitter because

Non-CLIL students	
230505	yes eh I thought it could be great
230510	would you eh where there is a hotel with a big pool eh the (X) are beautiful {we are} we can see the: sea eh ... for the breakfast if it was really for me I would choose neither because I don't like hot weather at all
230503	it's would be good a mix of the

5.3.4.3. Subordinate clauses:

Figure 17: Number of clauses per AS-units in both groups



As we can see from this graph, the number of clauses among CLIL students exceeds that of non-CLIL students. Beyond observing that in terms of grammatical complexity, CLIL students seemed to be more advanced, it was noticed that they have a greater tendency to produce subordinate sentences. This adds weight to the grammatical complexity of CLIL students. However, it should be noted that this difference is minor and therefore not fundamentally significant.

5.3.5. Conclusion

In order to answer the hypothesis concerning the complexity of the two groups, it can be argued that globally it has been observed that the creativity and therefore the complexity of the structures of CLIL students are more advanced than the complexity of the non-CLIL learners. In terms of grammatical complexity, it has been observed through numbers and through personal observations that their productions seem to be more complex. It is likely, although not proven, that this is explained because these students are more stimulated in an environment that promotes proper cognitive development. However, it should be pointed out

that the difference between the two groups was not striking either. This variable should therefore be thought of as the fact that CLIL learners seem to have more developed interlanguage skills because they use more complex grammatical structures, but their language system still needs to be modified, perhaps involving, for example, the explanation of certain rules in order to improve knowledge that they have only partially assimilated.

Second, the hypothesis that the lexical complexity of CLIL students would be more developed than that of non-CLIL students is refuted. Indeed, no major differences were observed in terms of lexicon. On the contrary, it has been proven that in terms of figures the non-CLILs slightly outnumbered their CLIL colleagues. This is probably due to the context imposed by the topic

CONCLUSION

This MA dissertation aimed to evaluate the level of acquisition of CLIL and non-CLIL learners' speaking skill in French-speaking Belgium using three essential variables for judging their proficiency, which were complexity, accuracy, and fluency. As aforementioned, these variables have been considered significant over the years due to changes in vision within fields such as Second Language Acquisition and have become essential factors in intelligible judgments of learners' proficiency. Through the various theoretical chapters, it has been possible to discuss strengths and weaknesses that this work might have encountered in its practical investigations.

In the first chapter, it has been shown that the CLIL teaching program promotes education that combines culture, knowledge, language skills and cognitive development (Dalton-Puffer 2008: 141). This type of teaching is promoted at both international and European level, since the interest in developing multilingualism is constantly growing. Within the Belgian State, it was noted that deep linguistic tensions rooted in national identity justified the attempts to harmonize the linguistic landscape. By examining studies on the subject, it was noted that the main contributions of this type of teaching are the acquisition of vocabulary, fluency, the development of creativity as well as risk-taking behaviors in learning and therefore possibly faster learning (see for instance Ibid. 144, Naiman 1995, Mewald 2004, Rieder & Hüttner 2007, as cited in Ibid.). However, this type of teaching did not seem to be as effective as the Canadian immersive teaching, which is the source of inspiration of these dual immersive teaching systems. Indeed, it seems that CLIL instruction does not provide as much/enough contact with the target language (Day and Shapson 1996, as cited in Ibid.). It is on this foundation that the reflection has been carried out and continued in the following chapter.

The second chapter has provided insights into key phenomena in SLA studies. It was important to highlight these phenomena in order to be as complete as possible in the research that was to be conducted. The third chapter addressed the limitations and strengths of the variables that were to be used to conduct the practical analysis. On the basis of these reflections, different visions such as the psycholinguistic approach to cognition and the construction of knowledge were presented.

Subsequently, several research questions and hypotheses based on the scientific literature were formulated. Through a practical case study, it was possible to draw a number of conclusions. Notably, in terms of complexity, it appears that the creativity and complexity of the structures of CLIL learners seem to be more developed. This conclusion was drawn as a result of certain observations. In particular, it was noted that the use of complex tenses was higher in CLIL productions than in non-CLIL productions. A second observation was the slightly higher ratio of subordinate clauses in CLIL productions. The use of modal verbs was also mainly observed in CLIL productions. However, while it was initially argued that the vocabulary acquisition of these students would be more developed than that of non-CLIL students, this analysis refuted this assumption because no significant difference was observed between the two groups. The slight difference observed tended to be in favor of the non-CLIL students. It was therefore suggested that if CLIL students had been observed in the previous scientific literature to outperform their non-CLIL counterparts in terms of vocabulary acquisition, this could be due to the fact that they are in contact with more specific vocabulary because of the content of their compulsory classes. Non-CLIL students do rarely encounter such technical (disciplinary/content specific) vocabulary as content courses are mainly taught in their mother tongue. Furthermore, the data covers conversations on a given topic, which justifies that the vocabulary found in the dialogues is essentially centered on the same lexical field since it deals with the same subject.

Regarding the second variable, accuracy, it was observed that both CLIL and non-CLIL students did not produce error-free speech. However, since it seems that CLIL students use more complex structures, it might seem logical that these structures are not perfectly integrated in their interlanguage. Moreover, these students are not or hardly ever confronted with grammar focused instruction in their CLIL courses. This could explain why they tend to use more complex structures encountered through input in their CLIL courses but have no specific guidance as how these accurately.

Regarding the third variable, fluency, it was observed that overall CLIL students seemed slightly more fluent than non-CLIL students. However, the difference is not significant as it is solely justified by the fact that some measures such as the number of filled or silent pauses are slightly smaller than in non-CLIL productions. However, CLIL students tend to produce more words over a roughly equal time period than non-CLIL students. In order to assess fluency, it was decided to use a more subjective analysis to determine whether the overall impression

seemed to be more fluent among CLIL or non-CLIL students. Unfortunately, due to the limited access to audio data, it was not possible to listen to each conversation. Based on a purely subjective judgment using my experience as a master student specializing in languages and didactics, the conclusion was drawn that overall CLIL students seemed to be more fluent than non-CLIL students based on partial listening and analysis of the transcripts.

The practical investigations have therefore made possible to come to partial conclusions, but it must be pointed out again that these have been limited by a number of factors. One of the main limitations of this work is the size of the corpus. It is partly representative of a population of CLIL and non-CLIL students since the data was extracted in the framework of a large scale-interdisciplinary project. However, due to the 2020 health crisis of covid-19, it was necessary to restrict certain investigations. In particular, because of the restricted access to certain documents in the corpus such as audio files. Moreover, regarding the scientific literature, it is important to specify that all of it was extracted from the Internet thanks to the university's online resources.

On the basis of what has been observed in the practical analysis, it must be admitted that CLIL education seems to have positive consequences on areas such as the cognitive development of the students. As far as language skills are concerned, it appears that these skills are inclined to be developed but that they may not reach a significant level of improvement compared to that of non-CLIL learners. The observed subjective example of grammar mistakes can be mentioned. These errors are one of the main shortcomings observed as a consequence of CLIL teaching. It is possible that this lack of theoretical structure contributes to the lack of error-free language development. However, if communicative competence was to be characterized as successful communication in another language, not all of the errors observed in the corpus severely hinder comprehension. It is therefore necessary to review the standards and to claim that even if the students make mistakes, if they manage to communicate in the target language and make themselves understood, it can be considered as successful communication. We must therefore consider this distinction between what is expected and what is achieved. In purely normative terms, both groups of students make mistakes that are more or less significant at advanced level since they are considered as basic (e.g. use of singular forms instead of plural forms). However, in terms of communicative capacity it seems that these errors do not prevent the fundamental understanding of the message. It is therefore interesting to position oneself as to whether one wants the priority to be the transmission of an understandable

message or the normative quality of the language communicated in order not to interfere with comprehension.

In order to conclude this paper, it is therefore argued on the basis of the practical observations that some gains are achieved through CLIL teaching, but that this system could be reviewed in order to improve, for instance, skills such as grammatical accuracy in learners' productions. It is therefore important to consider if one wishes to improve this type of education, the goals sought by the system and whether communicative competence is defined in a normative way or in terms of communicative success.

BIBLIOGRAPHY

Adrián, M. M., & Mangado, M. J. G. (2015). L1 Use, Lexical Richness, Accuracy and Syntactic Complexity in the Oral Production of CLIL and NON-CLIL Learners of English. *Atlantis,37*(2). 175-197. Retrieved from: https://www.jstor.org/stable/24757789

Akdemir, A. S. (2016). Willingness to Communicate (WTC) in L2: An Affective Construct of Language Learning Process. *Journal of Graduate School of Social Sciences*, *20*(3). 839-855. Retrieved from: https://dergipark.org.tr/en/download/article-file/265486.

Beheydt, L. (1993). Second language learning in Belgium. In: Ager, D., Muskens, G., & Wright, S. (Eds.), *Language education for intercultural communication* (pp.15-22). Clevedon: Multilingual Matters LTD.

Beheydt, L. (1995). The linguistic situation in the New Belgium. In: Wright, S., & Kelly, H. (Eds.), *Language in contact and conflict: Contrasting experiences in the Netherlands and Belgium* (pp.48-64). Clevedon: Multilingual Matters LTD.

Beheydt, L. (2001). *Tweetalige scholen: vloek of zegen?* In: W. Vandeweghe, Evenepoel S., Maes A., De Meersman A. (Eds.), *Polyfonie. Opstellen voor Paul van Hauwermeiren* (pp.13-21). Gent : Mercator Hoogeschool.

Bui, G., & Skehan, P. (2018). Complexity, accuracy, and fluency. In: Liontas J.l. & TESOL International Association, *The TESOL encyclopedia of English language teaching* (pp.1-7). John Wiley & Sons. DOI: 10.1002/9781118784235.eelt0046.

Bulon, A., Hendrikx, I., Meunier, F., & Van Goethem, K. (2017). Using global complexity measures to assess second language proficiency: Comparing CLIL and non-CLIL learners of English and Dutch in French-speaking Belgium. *Papers of the Linguistic Society of Belgium*. *11* (1). 1-25. Retrieved from: http://hdl.handle.net/2078.1/184992

Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In: Housen, A., Kuiken, F., & Vedder, I. (Eds.), *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp.21-46). Amsterdam: John Benjamins Publishing.

Bulté, B., & Housen, A. (2015). Evaluating short-term changes in L2 complexity development. *Círculo de Lingüística Aplicada a la Comunicación, 63*. 42-76. DOI : 10.5209/rev_CLAC.2015.v63.50169.

Bulté, B., & Roothooft, H. (2020). Investigating the interrelationship between rated L2 proficiency and linguistic complexity in L2 speech. *System, 91*. 1-15. DOI: 10.1016/j.system.2020.102246

Chopey-Paquet, M. (2008). CLIL in French-speaking Belgium. Transforming paradox into potential. In: CN. Coonan, (Eds.*). CLIL e l'apprendimento delle lingue. Le sfide del nuovo ambiente di apprendimento* (pp.239-247). Venezia: Libreria Editrice Cafoscarina.

Chopey-Paquet, M., & Amory-Bya, N. (2007). Mission CLIL-possible: The quest to generate and support good CLIL practice in francophone Belgian state comprehensive catholic secondary schools. In: D. Wolff & D. Marsh. *Diverse contexts–converging goals. CLIL in Europe.* Frankfurt: Peter Lang. 70-90. Retrieved from: https://www.academia.edu/download/31271661/ChopeyPaquet_Bya_2006.pdf.

Dalton-Puffer, C. (2008). Outcomes and processes in Content and Language Integrated Learning (CLIL): Current research from Europe. In: W. Delanoy & L. Volkmann (Eds.), *Future perspectives for English language teaching* (pp.139-157). Heidelberg: Carl Winter. Retrieved from: http://www.academia.edu/download/36790465/CLIL_research_overview_article.pdf.

Darn, S. (2006). Content and Language Integrated Learning (CLIL): A European Overview. Retrieved from: https://eric.ed.gov/?id=ED490775

De Smet, A., Mettewie, L., Galand, B., Hiligsmann, P., & Van Mensel, L. (2018). Classroom Anxiety and Enjoyment in CLIL and Non-CLIL: Does the Target Language Matter?. *Studies in Second Language Learning and Teaching*, *8* (1), 47-71.

Dewaele, J. M. (2000). Saisir l'insaisissable? Les mesures de longueur d'énoncés en linguistique appliquée. *International Review of Applied Linguistics in Language Teaching (IRAL)*, *38*(1), 17-34. DOI: 10.1515/iral.2000.38.1.17

Dewaele, J. M. (2009). Individual differences in second language acquisition. In: W.C. Ritchie, & T.K. Bhatia (Eds.), *The new handbook of second language acquisition 2* (pp.623-646). Emerald Group Publishing Limited.

Drozdzial-Szelest, K., & Pawlak, M. (Eds.). (2014). *Psycholinguistic and Sociolinguistic Perspectives on Second Language Learning and Teaching: Studies in Honor of Waldemar Marton*. Springer Science & Business Media.

Ellis, R., & Barkhuizen, G. P. (2005). *Analysing learner language*. Oxford: Oxford University Press.

Eurydice. European Commission. (2006). *Content and Language Integrated Learning (CLIL) at School in Europe.* Brussels: Eurydice.

Foster, P., Tonkyn, A., & Wigglesworth, G. (2000). Measuring spoken language: A unit for all reasons. *Applied linguistics*, *21*(3), 354-375. DOI: 10.1093/applin/21.3.354

Gass, S. M. & Selinker, L. (2013). *Second language acquisition: An introductory course*. New-York: Routledge.

Genesee, F. (1998). French immersion in Canada. In: Edwards, J. (Eds.), *Language in Canada.* Cambridge: Cambridge University Press.

Genesee, F., & Jared, D. (2008). Literacy development in early French immersion programs. *Canadian Psychology/Psychologie canadienne*, *49*(2), 140-147. DOI : 10.1037/0708-5591.49.2.140.

Gilquin, N., & Granger, S. (2015). Learner language. In: Biber, D., & In Reppen, R. (Eds.) *The Cambridge handbook of English corpus linguistics* (pp.418-435). Cambridge: Cambridge University Press.

Hendrikx, I., Van Goethem, K., & Wulff, S. (2019). Intensifying constructions in French-speaking L2 learners of English and Dutch: cross-linguistic influence and exposure effects. *International Journal of Learner Corpus Research*, 5(1), 63-103. DOI: 10.1075/ijlcr.18002.hen.

Hiligsmann, P. & Van Mensel, L. & Galand, B. & Mettewie, L. & Meunier, F. & Szmalec, A. & Van Goethem, K. & Bulon, A. & De Smet, A. & Hendrikx, I. & Simonis, M. (2017). Assessing Content and Language Integrated Learning (CLIL) in French-speaking Belgium: linguistic, cognitive, and educational perspectives. *Les Cahiers de recherche du Girsef*. 1-24.

Housen, A., Kuiken, F., & Vedder, I. (Eds.). (2012). *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (Vol 32). Amsterdam: John Benjamins Publishing.

Kormos, J. (2011). *Speech production and second language acquisition*. Routledge.

Lazaruk, W. (2007). Linguistic, Academic and Cognitive Benefits of French Immersion. *Canadian Modern Language Review, 63*(5), 605-628. Toronto: University of Toronto. DOI: 10.3138/cmlr.63.5.605.

Long, M. H. (2017). *Problems in second language acquisition*. Routledge.

Marsh, D. (2013). *Content and Language Integrated Learning (CLIL). A Development Trajectory* (Doctoral dissertation). Universidad de Córdoba. Retrieved from: http://hdl.handle.net/10396/8689

Mauranen, A., & Ranta, E. (2009). *English as a lingua franca: Studies and findings*. Cambridge Scholars Publishing.

Mota, M. (2003). Working memory capacity and fluency, accuracy, complexity, and lexical density in L2 speech production. *Revista Fragmentos*, 24, 69-104. DOI: 10.5007/fragmentos.v24i0.7659.

Michel, M. (2017) Complexity, Accuracy and Fluency (CAF). In: S. Loewen & M. Sato *The Routledge Handbook of Instructed Second Language Acquisition* (pp.50-68). London: Routledge.

Michel, M. C., Kuiken, F., & Vedder, I. (2007). The influence of complexity in monologic versus dialogic tasks in Dutch L2. *International Review of Applied Linguistics in Language Teaching*, *45*(3), 241-259. DOI: 10.1515/iral.2007.011

Rieger, C. L. (2001). Idiosyncratic fillers in the speech of bilinguals. *ISCA Tutorial and Research Workshop (ITRW) on Disfluency in Spontaneous Speech*, 81-84.

Rustipa, K. (2011). Contrastive analysis, error analysis, interlanguage and the implication to language teaching. *Ragam Jurnal Pengembangan Humaniora*, *11*(1), 16-22. Retrieved from: https://www.academia.edu/download/34284304/contrastive_and_error_analysis.pdf

Saville-Troike, M. (2012). *Introducing second language acquisition* (2ème éd.). Cambridge University Press.

Schoonjans, E. (2012). A contextual perspective on oral L2 fluency. *Eurosla Yearbook, 12*(1), 135-163. DOI: 10.1075/eurosla.12.08sch.

Shumin, K. (2002). Factors to consider: Developing adult EFL students' speaking abilities. In: J.C. Richards & W. A. Renandya *Methodology in language teaching: An anthology of current practice* (pp.204-211). Cambridge University Press.

Skehan, P. (2003). Task-based instruction. *Language teaching*, *36* (1), 1-14. DOI: 10.1017/S026144480200188X.

Song, S. (2018). Second language acquisition theories. In S. Song, *Second Language Acquisition as a Mode-Switching Process* (pp.9-36). Palgrave Pivot, London.

Sung, K. Y. (2010). Promoting Communicative Language Learning through Communicative Tasks. *Journal of Language Teaching & Research, 1*(5), 704-713. DOI: 10.4304/jltr.1.5.704-713.

Towell, R. (2012). Complexity, accuracy and fluency from the perspective of psycholinguistic second language acquisition research. In: Housen, A., Kuiken, F., & Vedder, I. (Eds.). (2012). *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA* (pp.47-70). Amsterdam: John Benjamins Publishing.

Van Mensel, L., Hiligsmann, P., Mettewie, L., & Galand, B. (2020). CLIL, an elitist language learning approach? A background analysis of English and Dutch CLIL pupils in French-speaking Belgium. *Language, Culture and Curriculum*, *33*(1), 1-14.

Van Moere, A. (2012). A psycholinguistic approach to oral language assessment. *Language Testing, 29* (3), 325-344. DOI: 10.1177/0265532211424478.

Willemyns, R. (2002). The dutch-french language border in Belgium. *Journal of Multilingual and Multicultural Development*, *23*(1-2), 36-49.

Xie, F., & Jiang, X.M. (2007). Error analysis and the EFL classroom teaching. *US-China education review* 4(9), 10-14.

APPENDICES

1) Non-CLIL Corpus annotated:

a) Conversation 1 (Student 230508 and 230505)

<230508>| Hello | |how are you ? |</230508>

<230505> |fine and you ? |</230505>

<230508>| Oh fine| </230508>

<230505>| mm what kind of trip did you win ? | </230505>

<230508> |Ah mm I won eh a trip {to} mm . eh to Italy and you ? |</230508>

<230505> |eh Thailand| </230505>

<230508> |Oh that's cool| </230508>

<230505> |yes eh I thought :: it could be great .:: if you come with me eh in my trip | </230505>

<230508> |Oh yes but eh maybe my trip is better| </230508>

<230505> |why ? |<laugh> </230505>

<230508> |Oh eh because we can do: mm . eh . eh camping ||we can go eh .. :: to see the landscape eh | |we can eh walk around the landscape | |and if you like the barbecues ::we can do a lot of barbecues| <laugh> </230508>

<230505>| Oh yes but that's a lot of sports . || I don't really like sports eh |but eh my trip is a trip to relax| | . eh we can go to the beach|| mm have P(0,3) eh .. mm a big room in a: very cool hotel |and eh relax a lot :: cause there's a swimming pool in the[i:] hotel ||so: . I think my trip is a little bit better |<laugh> </230505>

<230508> |yes eh so you right |.| I can't say :: {that's} that is not perfect {to} to relax and be cool eh :: cause it's eh the holiday | |and yeah with the holiday :: you must be cool relax no stress P (0,1) . mm|| but . where eh will we go so: ? |</230508>

<230505>| where | |eh I don't know . mm |</230505>

<R> maybe you can also ask a little bit more what's on the pictures ? </R>

<230505>| yes mm what else can we do eh in Italy ? |</230505>

<230508> |eh we can do a tour of the Italy with s/ eh . bicycle eh| |and eh we can eh take .. mm . some things ::to eh sleep eh in the nature eh {at the} . in the night with the stars and| eh </230508>

<230505> |yes that sounds cool eh| </230505>

<230508> |mm Okay so it's like you want | | eh I love to travel | |so eh I will follow you|

</230508>

<230505> |Okay| . |so: you go to Thailand with me ? |</230505>

<230508> |Oh yes |. |it will be cool |.| I enjoy it| <230508>

<R> and usually to what kind of destination would you go ? with you parents . your friends where do you normally go ? </R>

<230508> <overlap /> |France <laugh> yes |</230508>

<230505> <overlap />| France| </230505>

<R> both of you ? </R>

<230508> |ah yes| </230508>

b) Conversation 2 (Student 230501 and 230504)

<230501> | hi | </230501>

<230504> | How are you today? | </230504>

<230501>| fine and you ? |</230501>

<230504> |fine you know next week there is {a } a party | |and I want to .. P(0,2) come with me ? | </230504>

<230501> |yes eh but I have eh also a party |.| and eh {I } I wanted eh ::that you come with me too |. mm <laugh> eh it eh </230501>

<230504> |{why eh} why do you want eh :: I come with you ? | </230504>

<230501> |eh because eh you are my: best friend | |so I want to go at this party with my best friend |</230501>

<230504> | and why eh your party is better than mine ? | </230504>

<230501> | eh because eh there will be eh pizza . | |and eh we will eh sing mm { a lot mm} a lot of 80's { musics } music| |and eh we will eh watch movie |</230501>

<230504> |okay .. | |but my party eh. it's also good :: because mm { it will be eh} it will have eh fireworks .. eh a lot of champagne and mm a lot of people . one thousand eh :: a friend eh told me . | | and eh we can come eh in a big car . in a limousine :: {I don't . and mm}. | | I don't know |.| do you like my eh (X) ? </230504>

<230501> |eh yes I like ::because eh I prefer: mine| </230501>

<230504> <laugh> |do you like {mine} my party ? {my party do |y/} </230504>

<230501> eh <laugh>| yes I like but| eh </230501>

<230504> |you prefer yours okay| </230504>

<230501>| {I want} I want to go eh . in my . | |and eh we will eh play . mm Monopoly

. and eh . and eh other games | (laugh) </230501>

<230504> |yes but mine my party eh there will have a lot of <foreign> alcool </foreign> beer and eh other |.| so: I don't know eh :: if it will have eh in your party alcohol | |but eh mine's probably better | <laugh> </230504>

<230501> {yes eh it can P(0,2).. no <laugh> mm } |yes if you want eh I can eh bring eh alcohol |
</230501>

<230504> |oh okay |. |at your party ? | </230504>

<230501>| yes |</230501>

<230504> |ah ok and eh| </230504>

<230501> | and a lot of pizza of chips of eh |</230501>

<230504> | and where is it ? |</230504>

<230501> | it is eh . eh at eh my neighbour's house| |so you can sleep| </230501>

<230504> |at your home ? | </230504>

<230501>| at my home yes| </230501>

<230504>| okay but my party is mm . next my home| |so you can also sleep in my . house |</230504>

<230501> |and eh if you come at my party { you: .. } :: you: { haven't . } haven't { to .. to . }P(>0,3) <foreign> <starts laughing> je sais plus comment on dit <stops laughing> </foreign> mm to bring money because eh . everything is eh { is} eh … </230501>

<230504> |free ? |</230504>

<230501> free |<laugh> </230501>

<230504>| okay . but eh you know:: { all the team. } whole team eh come to my party |. |and eh so: it's better . eh :: that you come to my party |<230504>

<230501> |yes but at my party there will eh be eh Maria| |and I know that you love Maria |<laugh>
</230501>

<230504> <laugh> eh .. P(>0,3) </230504>

<230501> |so you have to come at my party |</230501>

<230504> |okay I: will ask to: the team . to come at your party . next week| </230504>

<230501>| oh thank you . mm . | P (0,2) | |and eh me I will ask to my mother :: if you can sleep eh at eh my house |</230501>

<230504> |you must say| </230504>

<230501> | {she will be yeah} she will say yes| </230501>

<R> is it a birthday party ? </R>

<230501> | me yes |<laugh> </230501>

<230504> |not mine it's a: . simple . party |</230504>

<230501> |it's eh the birthday of eh { of } eh .. </230501>

<230504> |a friends| <laugh> </230504>

<230501> of my friends :: who is my neighbours | | { and } P (0,2) eh . and eh she has { eh twelve eh <foreign> non twelve c'est en néerlandais </foreign> eh eighty . } eighteen eh… </230501>

<230504> |years old| </230504>

<230501> years old | </230501>

c) Conversation 3 (Student 240303 and 240410)

<240303> |Hello| <first name of the partner>| how are you ? | </240303>

<240410> |I'm fine and you ? | </240410>

<240303> | eh fine (eh) | | what are you doing tonight ? | </240303>

<240410> mm ... P(0,3) |I'm invited eh .. in a party . {with} eh ss with a friend | </240410>

<240303> eh what eh .. |" {what .. will you} . what will you do <starts laughing> at this party |<stops laughing> </240303>

<240410> mm .. mm .. |{they} they are eh .. P(0,3) <laugh> music and fireworks {cham} / champagne |
</240410>

<240303> (mm) {co/} | come with me please :: because eh we will eat a pizza :: and play eh Monopoly |: | and after:: we will eh see eh {a[ei] film a} a[ei] movie |</240303>

<240410> |oh oh no | |I prefer my party because eh {the/} there is a <foreign> limousine | </foreign></240410>

<240303> | the <foreign> limousine </foreign> | | but if you come with me :: {we will eh} we will do a[ei]
<foreign> karaoké </foreign> </240303>

<240410> eh .. | I . don't . sing | </240410>

<240303> | you don't sing :: cause {you}: you don't have eh a[ei] good voice | or <laugh> but
</240303>

<240410> | it is horrible | </240410>

<240303> | but {eh pizza}: you know pizza is | eh </240303>

<240410> | I like pizza | | but I prefer <foreign> champagne | </foreign> <240410>

<240303> ah . {but eh ... P(>0,3) we: } |{what } what is your favourite movie ? |
</240303>

<240410> ouf eh | it's the[i:] Holiday with Cameron Diaz | </240410>

<240303> oh so eh | we will see this film eh tonight| | so come with me | <laugh>
</240303>

<240410> no eh P(>0,3). { I }. <sigh> . I . <foreign> { pfou je sais pas comment on dit ça . mais je l'ai déjà vu }
</foreign> </240410>

<R> I have already seen the movie </R>

<240410> <foreign> ouais bon </foreign> <laugh> mm .. eh I .. eh </240410>

<240303> | {you} you have already see it or not ? |. |.yes |.</240303>

<240410> no mm | I prefer eh dancing eh ..:: because it's eh |

<240303> | but if you want {after} after {the: }. the movie and eh monopoly party ::
we will go eh {in}. eh in a discotheque to dance | </240303>
<240410> | after | </240410>

<240303> | yes it will be fun | </240303>

<240410> | I prefer | </240410>

<R> who invited you to the party ? </R>

<240410> | mm in first P(>0,3). eh her party and after my party | </240410>

<R> yes I know but who invited you ? is it someone from school is it a friend ? </R>

<240303> |ah yes mm | |and who invited you ? | </240303>

<240410> | | ah mm my best friend | | </240410>

<240303> | you best friends ? | </240303>

<240410> | yeah { it is your it's your } it's eh her { birthday } birthday | (correcting
pronunciation) </240410>

<240303> mmm[2] |it's my best friend too| </240303>

<240410> |no it's impossible| <laugh> </240410>

<240303> | but I receive an invitation | | but {it }. it isn't at her birthday party | |it's eh
pyjama party |
</240303>

<240410> oh … <starts laughing> | it's boring | <stops laughing> </240410>

<240303> |no we will talk about . boys: and eh and girl :: that eh we ate |</240303>

<240410> <foreign> mais </foreign> | I prefer boys | </240410>

<240303> aha |so: you need to come with me . :: if you want to talk about boys your crush | (m)
</240303>

<240410> | { that's } that's { more } more great with boys | </240410>

<240303> aaaaaah | you don't want a pyjama party with eh girls only | </240303>

<240410> | no { I I I } I don't like this | </240410>

<240303> <overlap/> but </240303>

<240303> | it is not a pyjama party with boy | | it is always with girls| |so we don't need boy | <laugh>
</240303>

<240410> <foreign> oui mais </foreign> |I don't like |<laugh> </240410>

<R> so which party do you choose . both parties or </R>

<240303> |pyjama party| </240303>

<R> ok </R>

<240303> |best party| <laugh> </240303>
<240410> |{both} both| </240410>

d) Conversation 4 (Student 230502 and 230510)

<230502> |Hello|: <name of the pupil> </230502>

<230510> |Hi |<name of the pupil> </230510>

<230502> |how are you ? | </230502>

<230510>| I'm fine thank you| | what about you ? | </230510>

<230502> |{Fi/} eh I'm fine| |eh I won a: city trip in <foreign> Italië |</foreign>
</230502>

<230510> |Oh nice| </230510>

<230502> |what sort of eh . city trip do you like ? |</230502>

<230510> |I like city trips:: that hallow me to see a lot of eh sceneries and eh monuments |</230510>

<230502> |Ah it's great| . | you do you like to visit ? | </230502>

<230510> |yes of course| </230510>

<230502> | yes . me I prefer: mm going to eh . eh . city trip where eh I can eh . do eh bicycle
<laugh> | |I eh like eh sports . | | mm I like eh camping with a big fire .. | |I like barbecue |.| do you like it? | </230502>

<230510>| yes but it's not really my thing | | I prefer eh white .. eh beach with white sand :: where we can can see the sea eh |</230510>

<230502> |yes it's great |</230502>

<230510> | with the sun shining ||.. so: eh I won a trip to Thailand |. |would you eh where there is a hotel with a big pool eh the (X) are beautiful {we are} we can see the: sea eh ... for the breakfast| | there is a big eh <foreign> buffet |</foreign>| {I think} and eh I think:: that you may be interested in going with me ? | </230510>

<230502> |yes eh the destination eh is great|| but eh I don't like ::when eh the: weather is eh very hot . | (mm) <laugh> |but you don't like eh doing sport in eh no in holiday ? | </230502>

<230510> |I don't like sport at all| </230510>

<230502> |Oh okay .. mm| </230502>

<R> which food would you prefer ? </R>

<230510>| I like exotic food :: that's not pretty common so I think eh it's eh| <230510>

<230502> |a new culture| . |yes |. |me I prefer sausages .. hamburger . beef |<laugh> </230502>

<230510> |but that's pretty common food | | I think so| </230510>

<230502> |yes . yes yes . || but {I'm} mm I'm okay {to}: to go in Thailand ::if you very don't like eh . going with me in <foreign> Italië |</foreign> <laugh> </230502>

<230510> |Ok thanks| <laugh> </230510>

<230502> <laugh> |okay |</230502>

<R> and usually to what kind of destination would you go ? </R>

<230502>| Oh its depends| </230502>

<R> but rather to Italy or rather to Thailand ? </R>

<230502> |mm I went to: China . with my parent eh ||but never in Thailand |<laugh> </230502>

<R> and you ? </R>

<230510> |if it was really for me :: I would choose neither :: because I don't like hot weather at all|
<laugh>| I can't stand the heat so |<laugh> </230510>

e) Conversation 5 (Student 230515 and 230503)

<230515> |what type of eh travel is it ? | </230515>

<230503> |mm it's in the south of Europe ||and there is a lot of sun | |and there is the sea: | | we can . go hiking | |and eh .. there is a lot of things :: to do actually | { there are } mm | it's really mm close to the nature|
. | you know :: there are lakes | |and eh {it's really } eh you are free ::to do what you want | it's not like there is an hotel :: and you have to go there and there :: and visit everything| . | it's more eh more you know like :: walking and biking and things like that . and you ? | </230503>

<230515> | and eh do you like eh walking {and } eh and eh see the nature | | {or} eh or you prefer eh to be lazy eh in the beach or | </230515>

<230503> | I prefer eh this | | I prefer to . just walk and watch eh everywhere and take pictures |.| you know like eh .. these things but not like not being stressed by eh like the time | | or I have to do that | or shit I have to go there and there is so much to do . to be cool and take your time | <laugh> | okay and what about your ? | </230503>

<230515>| me it's in a hotel <laugh> in Asia eh | | {and I}and I prefer this because eh it's in the sun eh | | it's more lazy {I pre/} I prefer eh take the sun <laugh> mm and eh it's by eh plane ||and eh I like plane |<laugh> mm … </230515>

<230503> | mm it's the food | </230503>

<230515> <foreign> ouais </foreign> | yes mm {what food} eh what food eh | </230515>

<230503> | are we going to eat ? | </230503>

<230515> <foreign> ah oui </foreign> </230515>

<230503> | eh more like barbecue | | but {we can} eh we don't have eh specific type of food |.| we can choose :: and eh do eh what we want | | but we have to think about eh the bag we can . take eh a lot of eh of food .:: because there are everything to put in your bag :: and you have to carry it | mm | what about your ? | | I mean eh ::. {is there}are there some activities {to}: to do ? | | or it is just to make pool . beach eh| </230503>

<230515> <starts laughing>| it's just swimming pool beach <stops laughing> mm maybe eh gymnastic or little activities | </230515>

<230503> | for me gymnastic is not eh holiday | <laugh> </230503>

<230515>| mm and eh my food is probably eh a lot of think | <laugh> </230515>

<230503> <laugh>| Asian food | | or is it like you know you go to mm Turkey ? | </230503>

<R> Turkey yes </R>

<230515> | {Tur/} Turkey okay | </230515>

<230503> | and you go there | | and there are like chicken and eh French fries | | and you're like . it is really what you eat in Turkey | | or it's just the food that eh you <foreign> fin= enfin </foreign> for tourists ? | | You can eat the real food of the way eh the place you go | <230503>

<230515> <foreign> ouais </foreign> | yeah but eh {in my} in my picture it's probably {a} a special food|
</230515>

<230503>| that's good | </230503>

<R> Ok so you can take a minute to decide where you will go in the end . which destination ? </R>

<230503> | So: I like Asia . really . | | but me it's Italy . so there is sun too| . </230503>

<230515> | Thailand me | </230515>

<230503>| But I like to go . to Thailand | | but . if we don't stay like everyday <overlap /> in the hotel
</230503>

<230515> <overlap /> | in the hotel yeah| </230515>

<230503> because I'll be mad at the end of the | </230503>

<230515> | you prefer visit | </230515>

<230503> | yes I like to discover| | it's really what the country is |. | If I want to go to a hotel I go to eh to the sea | </230503>

<230515> |yeah I think the same eh yes | |{I I }I prefer (A+) fin=enfin (A-) mm I like eh relax | |but eh I like too eh go to walking or | </230515>

<230503>| it's would be good eh a mix of the two . || So we go to Italy but in an hotel | |or we go to Thailand you know |</230503>

2) CLIL Corpus annotated:

a) Conversation 1 (Student 230402 and 230408)

<230402> | So <name of the pupil> where are you going on vacation? | </230402>

<230408> | I won {a:} a vacation ::for going to Italy |.| {it's} eh it seems to be a great place and . very sunlight | | so . you should come | </230408>

<230402> | yeah well | | I: just won a . trip to Thailand | | I think it's even better than Italy |.| there is sunshine too | </230402>

<230408> | and there is women with eh cock | </230408>

<230402>| there is what? | </230402>

<230408> | In Thailand there is eh change women | </230408>

<230402> | oh well eh | <laugh> | we just don't go to Bangkok| | and that's it |. | we'll stay in the[i:] hotel on the beach |.| no woman with cock there | </230402>

<230408> | but I think . Italy {is is near the }: is in Europe | so I think :: it's better {it's near the:} it's near Belgium so: the . | {we} we wouldn't be so: | . eh I'd say eh . eh <foreign> je ne sais même pas le dire en français |</foreign> </230408>

<230402> | homesick . homesick? | </230402>

<230408> |yeah homesick | </230408>

<230402> | well I think it's the point :: if you go on vacation | you don't want to stay near your country your job {your} your problems | | so: you go as far as possible ::. that's what Thailand is a good choice |
</230402>

<230408> | yeah I'm agree on that | | but . {the} the trip I've won is an adventure trip | | so . we're going {to make to} to bicycle | | and eh we have a tent to make camping eh with a fire during the night . and all that | | so I think it would be great and . should be good with eh the adventure trip {we have} I have
</230408>

<230402> |maybe| | but eh you know:: I'm not very good at sport :: that's not something I love | |and mm I'm just exhausted with all my jobs and the kids ||and I just want some eh time to: relax and enjoy || so: in Thailand we could eh . stand eh in the sun on the beach . just sunbathing or: . swimming at the swimming pool of the hotel | |there is a huge swimming pool . | |and eh the food is amazing in Thailand too|
</230402>

<230408> | yeah but . you know:: if you don't like sports :: you should make sport :: to get fitter because you know you're a big dick eh {big big} | so: you should make sport | |and here there is a: some walking in the mountains || so it is very great | |and there is also lake:: where you can swim {instead of the of} instead of the swimming pool| . | it is eh {more} more huge than your swimming pool | |so I think it's better |<foreign> allez </foreign> | in this lake there is also fishes | | so . you can eh you can fish them | |or you can eh also: swim with {the} the fish . | | so I think it's great
</230408>

<230402> |I don't like fish| | I mean eh there is a lot of bacteria's|| and eh I I'm afraid of them | |that's why a swimming pool is great | |it's very hygienic and eh| </230402>

<230408> | yeah | | but you know Thailand is {very} . very polluted so . thanks to the <X> many polician|
<?> |eh so I don't think :: it's a great thing for eh your health |</230408>

<230402>| well it depends on :: where you go| | and I mean eh next to the beach in a great hotel: |
</230402>

<230408>| but like for the lake and the fish | |it depends :: which lake it is | |{and} eh and if it is a good lake:: you don't have bacteria on in it |</230408>

<230402> | yes but I mean eh the water will be coler |. |in Thailand there water be will be warm |. |we see many fish and eh with great colors more than in Italy| </230402>

<230408>| no because in Italy it's during the summer | |so the lake :: which its water stays at the same place so the water {wi/ will} eh will warm | |and eh so it would be al <XX> like in Thailand | |so I don't think {the} the temperature of the water is a problem for |</230408>

<230402> | but I want to do scuba diving | |I mean in Thailand it would be amazing |.| you know all the fish the sea . with eh| </230402>

<230408> |yeah but you just say that :: you don't like sport | |so why do you want to: | </230408>

<230402> |scuba diving isn't a sport | |it's some eh kind of relaxation| {it's} </230402>

<230408> |but {you can make} you can bike like relaxation not eh as a sport |</230408>

<230402> |in the mountains it's hard there are many <X> and eh | </230402>

<230408> |no there is mountains | |but you can bike eh in the valley or something like that . not only {in the moutain} . in the[i:] mountains |</230408>

<230402>| but we could eh travel on a boat | |and eh see small islands | </230402>

<230408> | yeah that's good thing the small island :: because you discover new places | |but . I think Italy is very good place:: because . {you} you have eh {you} near the sea the Mediterranean sea| | eh this is a great sea :: which is very warm | | and eh there is . great waves | | so . this I like but in the eh Indian ocean ||. I don't {like} like the Malaysia airline :: which loose their train| | so maybe {we} you will {ta/} / find eh a piece of the airplane :: that disappear | | . so I don't think :: is a good idea | </230408>

<230402> |you're talking about safety | |but I mean {you you} you want us to sleep in a tent? | |To do camping? | | I you know:: I hate that| | but in Thailand {we'll have} we'll sleep in a beautiful hotel with a beautiful bedroom {a} a big TV screen and eh extraordinary meals | </230402>

<230408> |If you do that stay at home |</230408>

<R> get to a conclusion </R>

<230402> |well eh why don't we go one day in Thailand and one day in Italy? || Or one week in Thailand and then one week in Italy? | </230402>

<230408> |eh yeah we can ask the company eh to do |</230408>

<230402> |I'm sure ::they'll agree| </230402>

<230408> |I think ::that could be a great opportunity| </230408>

b) Conversation 2 (Student 230311 and 230313)

<230311>| hi | <name of the pupil> </230311>

<2 30313> | hi: | </230313>

<230311> | how are you ? | </230311>

<230313> | I'm fine and you ? | </230313>

<230311> | {fine} fine I've been invited {to a : to a : .. } to a party tonight | | eh it's a party eh next to the school (eh) . with eh lots of people and a swimming pool || and eh I thought you could come with me :: because {the invi/} the invitation is for two {per/} people so | : </230311>

<230313> | I don't know ::which activities are programmed ? | </230313>

<230311> | well there is a: . sort of (A+) tombola (A-) like eh .. with lots of gifts | . eh <laugh>| {there is a eh .. eh … well} P(0,3) there is the swimming pool | | so we could go eh swim | | eh there is a barbecue .. and eh |
</230311>

<230313> | that seems cool (eh) | | I've also been invited to a party | </230313>

<230311> |yeah| <laugh> </230311>

<230313> | but it seems better than yours | <laugh> </230313>

<230311> | why ? | </230311>

<230313> | because eh there's an outdoor cinema: | |and eh we will [h]eat pizza :: which I like <laugh> | |and we can also eh do a karaoke | |or play eh some mm little games with our friends | </230313>

<230311> | yes but is there a DJ to your party ?| </230311>

<230313> | mm I don't know| | it's not said in the invitation so: | </230313>

<230311> | because there is a famous DJ to mine | |and his songs are really good | |so: you might maybe prefer | </230311>

<230313> | what's his name ? | </230313>

<230311> | eh it's DJ Julie | <laugh> </230311>

<230313> <laugh> s/ mm .. P(0,3) | I think :: we could go to yours :: because eh I really love :: listening to DJ's music and also :: swimming | |so :: I think it could be great | </230313>

<230311> | yes but maybe yours is better :: because there are no pizzas in my party |so|: . like .. | I don't know | </230311>

<230313> | it's quite a difficult choice | </230313>

<230311> |yes| </230311>

<230313>| maybe we can go to both | </230313>

<230311> | well no it won't be possible because { they are } :: I think they are the same time | </230311>

<230313> | oh yeah| . | so . just choose one | </230313>
<230311> | eh {isn't there .. } P(0,1) aren't there some of your friends :: going to one of the two parties | | don't you know ? | </230311

<230313> | I really don't know |</230313>

<230311> | well then if you want :: we can go to yours | . if eh . | and we can bring back the DJ and eh { take him} take him to your party | </230311>

<230313> | it's OK for me |</230313>

<230311> |Okay that's it |. |and <laugh> maybe we could ask for: some more eh . activities for the party like eh . eh … like some video games and eh board games | |eh { you n/ } you see what I mean ? |</230311>

<230313> | eh yes | <starts laughing>| I see| <stops laughing> </230313>

<R> What kind of party is this ? </R>

<230313> |I already said it | .. | I said (eh) karaoke and games outdoor cinema so |</230313>

<R> is it for a birthday or another occasion ? </R>

<230313> | yes I think :: it's for a birthday |</230313>

<230311> | and whose birthday {it is} is it ? |</230311>

<230313> | eh it's Fanny's birthday| . | she is eighteen | </230313>

<230311> | Oh Fanny yes I know her |. | I think :: she won't mind if I come | </230311>

<230313> | so don't know: | </230313>

<230311> | and if you want:: I can come | | and take you like eh at {eh ni/ eh} seven o'clock PM |
</230311>

<230313> | yes that's kind . | | thank you .. and so: | | eh we see each other tonight | <laugh> </230313>

<230311> | yes okay |. | see you tonight | </230311>

c) Conversation 3 (Student 240113 and 240117)

<240113> |Hello <cough> Hello | |how are you ? |</240113>

<240117>| I'm fine and you ? |</240117>

<240113>| mm really super good :: because I just won {a} : a trip to Italy for two persons| </240113>

<240117>| Oh gosh me too except that me it's in Thailand |</240117>

<240113>| well were could we go ? |</240113>

<240117> |sorry| </240117>

<240113> |where do you wanna go ? | </240113>

<240117> |I prefer Thailand| |it's more hot| |we travel by plane |.| there is a huge swimming pool |. |there is the sea |. |there is good food| </240117>

<240113> |yeah I know | |but we could do camping in Italy | |that's really good :: I love camping| | and I know ::you love it too |</240113>

<240117> |it's a bit boring| |I prefer the sea {li/} | |do you know how Thai food is good |. |just think about it| </240117>

<240113>| I know Thai food is good | | but we could some barbecue in Italy | |and there is some beautiful lakes :: you don't have to go to the ocean | </240113>

<240117> |yeah ok should be good |</240117>

<240113> |yes . we could even do some bicycle | </240113>

<240117> |Ok I would enjoy it maybe |</240117>

<240113> |ok |</240113>

<R> can you maybe continue for example what kind of holidays you usually go to . just to talk a little bit more </R>

<240117> | but I prefer Thailand :: because usually I go on diving holiday| | and scuba diving is very more comfortable in Thailand than in Italy |</240117>

<240113> | I don't know I prefer Italy because it's more:: what I do usually | . |I . with my parents I usually on a: on trip like by foot | and I just go around the country and visit it |</240113>

<240117> |Oh it's quite cool |<overlap /> I </240117>

<240113> <overlap />| I mean :: I love Thailand but |</240113>

<240117> | I would maybe enjoy it | |but I{al/} I often went to Italy it's not the kind of real |<?> | I like | </240117>

<240113> <overlap /> well </240113>

<240117> |It's too much European style| </240117>

<240113> |you're right :: maybe we could go to Thailand |</240113>

<240117> |yeah ok |</240117>

d) Conversation 4 (Student 230414 and 230411)

<230414> | So <name of the pupil> I've eh won a kind of lottery eh | | and I {wont} eh I won trip in Thailand for two person | | . would you want eh to come with me ? | </230414>

<230411> | not really you're not too | | eh P(0,1). no <laugh> it doesn't eh | </230411>

<230414> | why not ? | </230414>

<230411> | mm just eh give me reasons to come with you| </230411

<230414> | Thailand is a: great really a great country| | you can eh ride on elephants' eh back for example| . | eh {the} the beach are eh wonderful| | eh you eat eh it's wonderful eh too| | and eh .. yeah I need someone to come with me| </230414>

<230411>| but the thing is :: that I won a trip to: Italy| | and there are eh many activities to do too and so: eh| </230411>

<230414> | {we've already} eh we've already been going to Italy last year so | </230414>

<230411>| not me | | . I didn't go to Italy | | . we can eh so it's in the mountains| and there is a[ei] | </230411>

<230414> <overlap /> | ah so we would go skiing ? | </230414>

<230411>| there are very eh beautiful landscapes et cetera | | and eh we can eat some eh barbecues | | eh go in eh tents | </230411>

<230414>| I love barbecues | </230414>

<230411>| eh and eh do {some} . some biking for example mountain bike yes | | and eh| </230411>

<230414> | and {would you what} . where would we:: I don't know sleep ? for example | </230414>

<230411> | eh in a tent | </230411>

<230414> | a tent ? | </230414>

<230411>| yes so: in a kind of camping so | </230411>

<230414> | yeah ok eh {I} eh {is it} is it eh a cold weather or ? | </230414>

<230411> |hot weather I would say |</230411>

<230414> |Yeah but in Thailand you know that's . far away | |that's eh . <overlap /> really beautiful |
</230414>

<230411> <overlap /> |yes that's true |</230411>

<230414>| that's a really beautiful country| | I I've already seen eh some pictures on the internet| | it's really eh beautiful |</230414>

<230411> |but you said that you've been | |eh ah no you didn't {sorry sorry not} sorry | |you've been to Italy not to Thailand| </230411>

<230414> | yeah {no no} no | </230414>

<230411>| I have to admit :: that Thailand is {a} very {interesting} eh an interesting country| .. mm |do you know some other activities:: that we can do: there ? | </230411>

<230414> |eh yeah maybe we could visit eh the: capital city the Bangkok| . | or eh I don't know maybe just lie down on the beach | |or eh I don't know| </230414>

<230411> |but {what is interesting in Bangkok ?} what is interesting in Bangkok? | </230411>

<230414> |I don't know:: I've never been to Bangkok yet| | but I think:: it could be nice {to} to visit some eh mm temples or things like that |</230414>

<230411>| but aren't you interested ::in eh looking at very beautiful landscapes in the mountains eh such as lakes | |or eh I don't know just mountains ? | </230411>

<230414> | that could be great too | |{I} I don't know it's quite difficult eh | </230414>

<230411> |But the thing is Thailand {is} is far away | |I have to admit :: that's better in Italy it's in Europe |.| it's a bit {like} eh like we have in Belgium not really the mountains etcetera but|</230411>

<230414>| yeah we could on {your} eh our own . not with eh your mm with eh a winning trip you know| </230414>

<230411> |yes that's true .. || but what's what would I do with this trip . ||so? I won it so | </230411>

<230414> |maybe you could sell it to an American for example | |that's . yeah |</230414>

<230411> <starts laughing> |why an American? || why ? |<stops laughing> <overlap /> |how could I|
</230411>

<230414> <overlap /> | yeah he lives far away|| so {he} he doesn't have the[i:] opportunity {to:} ::to come here and in Europe very often| |. so it could be great to: foreign| </230414>

<230411> | but{how would} how would I find this American ? | </230411>

<230414> |internet ? | <laugh> </230414>

<230411> <laugh> |yeah … mm| </230411>
<230414> |come on dude | <laugh> </230414>

<R> you can ask a little bit more what's on the pictures of the other person ? </R>

</230414>

<230411> |So what are {your/} on your pictures ? | </230411>

<230414> |eh I have the picture of mm a very nice swimming pool in a very nice hotel |</230414>

<230411> |how so it's in a pool {in a hotel} in an . hotel ? | if you're eh ah| </230411>

<230414> |yeah that's it | </230414>

<230411> |so I admit :: it's better than camping |<laugh> </230411>

<230414> | yeah camping is a bit rude | | and . I don't really like camping |.| eh on a eh on a . other picture eh there is plane| | so . I guess we'd go by plane {to}: to Thailand that's obvious| | I think {we can't} eh we can't go {by} eh by car or things like that . eh </230414>

<230411> | and so the plane is eh involved in {the: . the} :: what you've won ? | </230411>

<230414>| the price of the plane ? | | yeah . yeah I think yeah . the plane {the the} the price is eh . it's free too| | . so what do you have on your pictures ? | </230414>

<230411> | so I have some eh mountain bikes . a tent so {to:} to show camping| . |beautiful landscapes with mountains and eh flowers and eh all green stuff eh | |another . great landscape ::where you can see eh people {who} :: who walk I suppose | |so we can do a lot of walking {in} in Italy . | | and eh barbecue or kind of barbecue on {some} eh some eh . yes | </230411>

<R>so if you want you can conclude now </R>

<230414> | Okay so | </230414>

<230411> | so: mm on the one hand I just won this trip | |but I on the other than :: I think that Thailand is |
</230411>

<230414> | Thailand is great hé | </230414>

<230411> | yeah that's true :: I have to admit it | </230411>

<R> so you're going to Thailand? </R>

<230414> | yeah we are going to Thailand | </230414>

<230411> | yes I think :: that we'll go to Thailand | </230411>

e) Conversation 5 (Student 230312 and 230316)

<230312> | hi | <name of the pupil> </230312>

<230316> | hi | <name of the pupil> </230316>

<230312>| eh how are you ? | </230312>

<230316>| I'm fine and you ? | </230316>

<230312>| fine| </230312>

<230316> | mm I received an invitation for two persons :: to go to a big party mm in eh Miami
<laugh> this weekend on Saturday | | are you free ? | </230316>

<230312> eh | I also received two {invi/ } invitations for a party in Los Angeles . | | so: I wanted :: told to you about that | </230312>

<230316> | oh . and . what is the party about ? | </230316>

<230312> | eh it's a big party with a lot of alcohol and . not no drugs |<laugh> eh | but lot of people | and . we'll play: . a lot of things like eh alcohol games and karaoke and .. aqua pony | <laugh>
</230312>

<230316>| yeah that sound great | | but mine eh seems good too | mm eh there will be P(0,1) eh lots of people | eh it's a { fancy } a kind of fancy party eh with fireworks eh big light show and . champagne |
<laugh> | mm we can even go there eh with a limousine |. | it's free for us | | so: I think it might be cool |
<laugh> </230316>

<230312> |all right |. |that sound great |.| but eh we have " eh eh an outdoor cinema . eh with a lot of popcorn and . chips | </230312>

<230316> | oh yes why not | | but . I think the one in Miami can be great :: because there will be a lot of famous people|: | and we can meet them | | and go into the . VIP board and things like that |.| and we can also visit the city: eh just before | |and then stay for: . some days there| </230316>

<230312> | yeah that sound great | </230312>

<230316> | and eh .. I don't know |. | which one do you prefer ? | </230316>

<230312> |eh .. I really like mine | |but yours sounds better | </230312>

<230316> | oh so: |. I think :: we should go to . so: | </230316>

<230312> | yes | | so: . it's on Saturday ? | </230312>

<230316> | yes it's on Saturday | </230316>

<230312>| so: when do we need to go: ? where ? | </230312>

<230316> |eh we must eh go: this afternoon eh to Brussels | . | take the plane . | | { and eh there we will be . } when we we'll arrive to Miami there would be: a limousine to take us to . the party | | and yes . before we'll go { mm . din/ eh } to {a dinner } a fancy dinner and | </230316>

<230312> |okay| </230312>

<230312> |okay so: | |I'm going to prepare {my: } my luggage | </230312>

<230316> |okay me too |. |we'll meet . at the airport| </230316>

<230312> |yeah |</230312>

<R> do you know who sent the invitation ?</R>

<230316> | mm it's eh my uncle <laugh>:: he lives in Miami | |and he's an agent for: . big parties like . he: creates events ||and . he invited us . ||and {who } who eh invited you to your party ? </230316>

<230312> | {mm } my cousin . {she is} eh .. she is a big eh party maker <laugh> | | and eh she is very: eh famous in . the . parties {and} and everything| </230312>

<230316> | Oh great | </230316>

<R> will you meet new people there ? </R>

<230312> | my party is more eh like eh .. :: were you know everyone in the party and |</230312>

<230316> |{ yes} maybe yes this is a bigger party | | but we . { might . } eh . might not have the[i:]. possibility ::to get to know . people :: like we could in yours :: because it's eh bigger | |and . yours seems smaller | |so we could . talk more in yours | | but . I don't know :: it depends on :: what we want <laugh> |
</230316>

<230312> | I want to meet new people | </230312>

<230316> | okay so maybe we should go to yours | </230316>

<230312> | no to yours | <laugh> </230312>

<230316> | oh <laugh> |.| oh yes |. | I didn't understand what happen but ok |</230316>

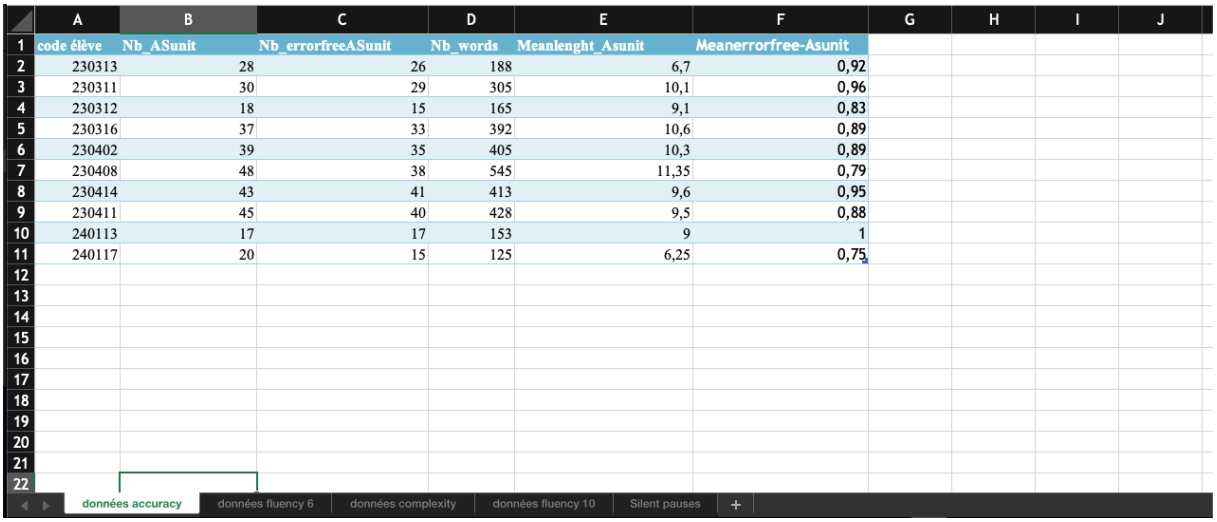
<R> ok so is it yours ? </R>

<230316>| yes okay | </230316>

<230312> | yes | </230312>

3) Excel files containing CLIL data

a) Overview of the entire table



	A	B	C	D	E	F	G	H	I	J
1	code élève	Nb_ASunit	Nb_errorfreeASunit	Nb_words	Meanlenght_Asunit	Meanerrorfree-Asunit				
2	230313	28	26	188	6,7	0,92				
3	230311	30	29	305	10,1	0,96				
4	230312	18	15	165	9,1	0,83				
5	230316	37	33	392	10,6	0,89				
6	230402	39	35	405	10,3	0,89				
7	230408	48	38	545	11,35	0,79				
8	230414	43	41	413	9,6	0,95				
9	230411	45	40	428	9,5	0,88				
10	240113	17	17	153	9	1				
11	240117	20	15	125	6,25	0,75				
12										
13										
14										
15										
16										
17										
18										
19										
20										
21										
22										

b) Accuracy data sheet

	A	B	C	D	E	F
1	code élève	Nb_ASunit	Nb_errorfreeASunit	Nb_words	Meanlenght_Asunit	Meanerrorfree-Asunit
2	230313	28	26	188	6,7	0,92
3	230311	30	29	305	10,1	0,96
4	230312	18	15	165	9,1	0,83
5	230316	37	33	392	10,6	0,89
6	230402	39	35	405	10,3	0,89
7	230408	48	38	545	11,35	0,79
8	230414	43	41	413	9,6	0,95
9	230411	45	40	428	9,5	0,88
10	240113	17	17	153	9	1
11	240117	20	15	125	6,25	0,75

c) Fluency data sheet (6 students)

	A	B	C	D	E	F	G
1	code élève	Nb_silentpauses	Nb_repetitions	Nb_rephrases	Nb_filledpauses	Nb_lexicalfillers	Nb_nonlexicalfillers
2	230313	6	0	0	15	0	15
3	230311	4	5	4	26	2	24
4	230312	1	3	1	19	1	18
5	230316	4	5	2	31	1	30
6	230408	0	16	5	30	2	28
7	230402	0	3	1	29	7	22

d) Complexity data sheet

	A	B	C	D
1	code élève	Nb_clause/ASunit	Lexical_density (content/total)	type/token ratio
2	230313	1,28	0,53	0,52
3	230311	1,26	0,46	0,39
4	230312	1,1	0,6	0,55
5	230316	1,32	0,38	0,49
6	230402	1,23	0,46	0,35
7	230408	1,37	0,42	0,28
8	230414	1,09	0,53	0,31
9	230411	1,2	0,51	0,31
10	240113	1,5	0,6	0,46
11	240117	1,1	0,59	0,58

e) Fluency data sheet (10 students)

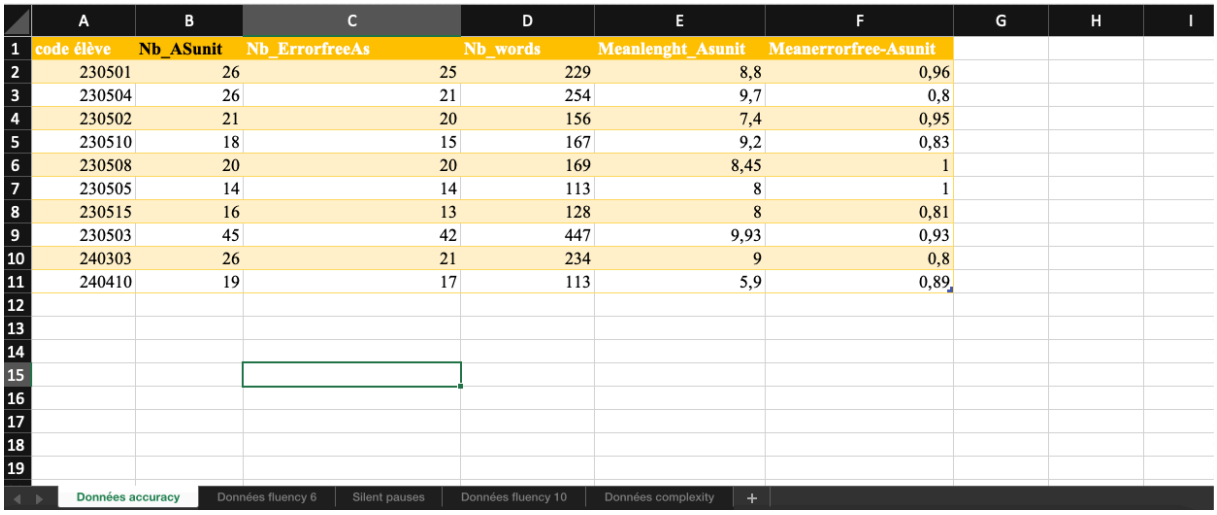
	A	B	C	D	E	F
1	code élève	Nb_repetitions	Nb_rephrases	Nb_filledpauses	Nb_lexicalfillers	Nb_nonlexicalfillers
2	230313	0	0	15	0	15
3	230311	5	4	26	2	24
4	230312	3	1	19	1	18
5	230316	5	2	31	1	30
6	230402	3	1	29	7	22
7	230408	16	5	30	2	28
8	230414	12	5	51	5	45
9	230411	11	4	34	2	32
10	240113	1	0	5	1	4
11	240117	1	1	5	1	4

f) Silent pauses data sheet (6 students)

	A	B	C	D	E
1	code élève	Nb_silentpauses	Number of silent pauses P(>0,3)	Number of silent pauses P(0,2)	Number of silent pauses P(0,1)
2	230313	3	1	0	0
3	230311	4	1	0	1
4	230312	0	0	0	0
5	230316	1	0	0	1
6	230408	0	0	0	0
7	230402	0	0	0	0

4) Excel files containing non-CLIL data

a) Overview of the entire table



	A	B	C	D	E	F	G	H	I
1	code élève	Nb_ASunit	Nb_ErrorfreeAs	Nb_words	Meanlenght_Asunit	Meanerrorfree-Asunit			
2	230501	26	25	229	8,8	0,96			
3	230504	26	21	254	9,7	0,8			
4	230502	21	20	156	7,4	0,95			
5	230510	18	15	167	9,2	0,83			
6	230508	20	20	169	8,45	1			
7	230505	14	14	113	8	1			
8	230515	16	13	128	8	0,81			
9	230503	45	42	447	9,93	0,93			
10	240303	26	21	234	9	0,8			
11	240410	19	17	113	5,9	0,89			
12									
13									
14									
15									
16									
17									
18									
19									

Données accuracy | Données fluency 6 | Silent pauses | Données fluency 10 | Données complexity | +

b) Accuracy data sheet

	A	B	C	D	E	F
1	code élève	Nb_ASunit	Nb_ErrorfreeAs	Nb_words	Meanlenght_Asunit	Meanerrorfree-Asunit
2	230501	26	25	229	8,8	0,96
3	230504	26	21	254	9,7	0,8
4	230502	21	20	156	7,4	0,95
5	230510	18	15	167	9,2	0,83
6	230508	20	20	169	8,45	1
7	230505	14	14	113	8	1
8	230515	16	13	128	8	0,81
9	230503	45	42	447	9,93	0,93
10	240303	26	21	234	9	0,8
11	240410	19	17	113	5,9	0,89

c) Fluency data sheet (6 students)

	A	B	C	D	E	F	G
1	code élève	NB_silentpauses	Nb_repetition	Nb_rephrases	Nb_filledpauses	Nb_lexicalfillers	Nb_nonlexicalfillers
2	230501	4	8	5	54	0	54
3	230504	2	2	5	27	0	27
4	240303	3	9	4	29	0	29
5	240410	12	10	2	27	0	27
6	230508	1	3	1	36	0	36
7	230505	2	0	0	18	0	18

d) Complexity data sheet

	A	B	C	D
1	code élève	Nb_clause/ASunit	Lexical density	type/token ratio
2	230501	1,26	0,61	0,37
3	230504	1,3	0,47	0,35
4	230502	1,1	0,63	0,41
5	230510	1,3	0,52	0,49
6	230508	1,15	0,53	0,41
7	230505	1,1	0,6	0,52
8	230515	1,06	0,56	0,46
9	230503	1,08	0,46	0,3
10	240303	1,15	0,49	0,36
11	240410	1	0,59	0,4

e) Fluency data sheet (10 students)

	A	B	C	D	E	F
1	code élève	Nb_filledpauses	Nb_lexicalfillers	Nb_nonlexicalfillers	Nb_repetition	Nb_rephrases
2	230501	54	0	54	8	4
3	230504	27	0	27	3	4
4	230502	26	0	26	2	1
5	230510	15	1	14	1	1
6	230508	36	0	36	3	1
7	230505	18	0	18	0	0
8	230515	31	0	31	6	1
9	230503	37	10	27	4	1
10	240303	29	0	29	9	4
11	240410	27	0	27	9	2

f) Silent pauses data sheet (6 students)

	A	B	C	D	E
1	code élève	Silent pauses	Number of silent pauses P(>0,3)	Number of silent pauses P(0,2)	Number of silent pauses P(0,1)
2	230501	9	1	3	0
3	230504	5	1	1	0
4	240303	3	1	0	0
5	240410	12	4	0	0
6	230508	1	0	0	1
7	230505	3	1	0	0

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
Faculté de philosophie, arts et lettres
Place Blaise Pascal, 1 bte L3.03.11, 1348 Louvain-la-Neuve, Belgique | **www.uclouvain.be/fial**