UCLouvain

FIAL
Un éventail de formations, une multitude d'horizons

Faculté de philosophie, arts et lettres

Exploring difficulties in simultaneous interpreting

Insights from the English-French Louvain Corpus of Student Simultaneous Interpretations

Auteur : Thomas Simon
Promoteur(s) : Marie-Aude Lefer
Lecteur(s) : Tania Biondi
Année académique 2018-2019
Intitulé du master et de la finalité : Master en interprétation, en finalité spécialisée

Exploring difficulties in simultaneous interpreting

Insights from the English-French Louvain Corpus of Student Simultaneous Interpretations

Thomas Simon
Année académique 2018-2019
Supervision : Marie-Aude LEFER

Mémoire présenté en vue
de l'obtention du diplôme de
Master en Interprétation de conférence

Remerciements

Ce mémoire est le fruit d'un travail considérable et je n'aurais pas pu y arriver sans les contributions de nombreuses personnes.

Tout d'abord, j'aimerais remercier Madame Marie-Aude Lefer, promotrice de ce mémoire, pour les nombreuses corrections, sa disponibilité et sa gentillesse. Son aide m'a été plus que précieuse pendant ce master et je l'en remercie infiniment.

Ensuite, j'aimerais exprimer ma gratitude à mes colocataires, Marianne Schoos et Margaux Vanbellinghen, qui m'ont soutenu et encouragé pendant l'écriture de ce mémoire.

I would like to express my heartfelt thanks to Alexa Alva, a friend of mine that I have met 6 years ago during my year abroad in the United States, for her interest in my work and her corrections.

Enfin, je voudrais adresser mes plus sincères remerciements aux personnes qui sont restées à mes côtés et sans qui un tel résultat n'aurait pu être atteint ;

À mes camarades de classe qui se sont intéressés à mon travail,

À mes parents, Vincent et Elisabeth, à ma sœur, Marie, et à mon frère, Mathieu, pour leur soutien bien que la situation fût bien souvent difficile à comprendre.

Table of contents

TABLE OF CONTENTS **5**
INTRODUCTION **7**
1 CORPUS-BASED INTERPRETING STUDIES **9**
1.1 EARLY HISTORY 9
1.2 CIS: A COTTAGE INDUSTRY 12
1.3 CORPUS COMPILATION 13
1.4 RECENT RESEARCH ORIENTATIONS IN CIS 18
1.5 INTERIM CONCLUSION AND FUTURE PROSPECTS 21
2 TEACHING SIMULTANEOUS INTERPRETING **9**
2.1 INTRODUCTION 23
2.2 EARLY HISTORY OF SIMULTANEOUS INTERPRETING TRAINING 24
2.3 GILE'S EFFORT MODEL OF SIMULTANEOUS INTERPRETATION 24
2.4 DIFFICULTIES OF SIMULTANEOUS INTERPRETING, WITH A SPECIAL FOCUS ON ENGLISH AS A SOURCE LANGUAGE 25
2.4.1 Numbers *26*
2.4.2 Proper names *28*
2.4.3 Complex noun phrases *30*
2.4.4 Single-word terms *31*
2.4.5 Culture-specific items *31*
2.4.6 Idioms *34*
2.4.7 Phrasal verbs *36*
2.5 RESEARCH QUESTIONS 37
3 DATA AND METHODOLOGY **39**
3.1 CORPUS COLLECTION 39
3.2 DATA TRANSCRIPTION 41
3.3 CORPUS DESCRIPTION 43
3.4 DATABASE 46
3.4.1 Source speeches: types of difficulties *46*
3.4.2 Student interpretations: taxonomy of rendition types *51*
4 RESULTS AND DISCUSSION **55**
4.1 GLOBAL FEATURES 55
4.1.1 Delivery rate *56*
4.1.2 Lexical density *59*
4.1.3 Expansion rate *64*
4.1.4 Interim conclusion *66*
4.2 DATA ANALYSIS 69
4.2.1 Introduction *69*
4.2.2 Quantitative analysis *69*
4.2.2.1. Complex noun phrases *70*
4.2.2.2. Single-word terms 78

4.2.2.3. Numbers……..79
4.2.2.4. Proper names…….81
4.2.2.5. Interim conclusion……..83
4.2.3 Longitudinal analysis 86
4.2.4 Case study: zero renditions 89
5 CONCLUSION 93
6 APPENDICES 95
6.1 APPENDIX 1: STUDIES ON INTERPRETING BASED ON AUTHENTIC CORPORA (SETTON, 2011)..... 95
6.2 APPENDIX 2: CONSENT FORMS 99
6.3 APPENDIX 3: TREETAGGER TAGSET DESCRIPTION (ENGLISH) 107
6.4 APPENDIX 4: TREETAGGER TAGSET DESCRIPTION (FRENCH)..... 110
7 BIBLIOGRAPHY 113

Introduction

As an interpreter trainee at the UCLouvain, I directly understood why conference interpreting is considered a hard task. It takes a lot to be a good interpreter and to achieve that goal as students need to master different skills, as much in consecutive interpreting and in simultaneous interpreting. These skills could be divided into two main categories: content related, and non-content related. For example, non-content related skills are good intonation/tone/volume, the ability to avoid disfluencies such as repetitions, self-corrections or silent pauses. I think it is really important to take all these aspects into consideration and to work on them during the master as they account for a part of the interpreter job. However, the skills I am most interested in are content related skills, especially transfers from one language to another. My two passive languages are English and Spanish, and I work from these two into French, my mother tongue. Working from English or Spanish into French are two different things as languages are constructed differently. I am personally more comfortable working from English into French, which is why I decided to focus on the English-French language pair for the present study.

As previously stated, transfers from one language to another has really captivated my interest. For this reason, I have chosen to focus on linguistic difficulties in simultaneous interpreting, which is the mode of interpreting I wanted to investigate as it is more challenging for me than consecutive interpreting. In addition, throughout my master at the Louvain School of Translation and Interpreting (LSTI), I have had the chance to follow courses on the use of corpora in the fields of translation and interpreting. In the recent years, an increasing number of corpus-based interpreting studies have been carried out, and I wanted to carry out my own corpus-based interpreting study. There have also been many studies focusing on interpreting quality assessment or teaching simultaneous interpreting, which are two aspects I also wanted to talk about in the present study.

This thesis focuses on the linguistic items that are potentially challenging for interpreter trainees (and professional interpreters) and aims to shed light on the most challenging linguistic item(s) and on the quality of students' performances across time. Chapter 1 presents an overview of corpus-based interpreting studies (CIS) and is divided into four main parts: the history of the field, the corpus compilation, recent research orientations in the field and the future prospects of the field. In Chapter 2, the focus is on teaching simultaneous interpreting (SI). The history of teaching SI is first discussed, Gile's Effort Model is then presented, followed with an analysis of seven difficult linguistic items in SI, with a special focus on English as a source language. This last part of the chapter is devoted to the two research questions. Chapter 3 presents the methodology used in the present study and gives details about the database. The analysis of three global features, namely delivery rate, expansion rate and lexical density, are presented in Chapter 4, as well as the results of the data analysis. The aim of this chapter is to be able to answer the two research questions. Finally, Chapter 5 is about the conclusions drawn from this study carried out on interpreter trainees and about the possible future research and pedagogical applications.

1 Corpus-based interpreting studies

In this first chapter, I focus on different aspects of corpus-based interpreting studies (CIS). Section 1.1 focuses on the early history of the field. In Section 1.2, I explain why CIS is still a field in its infancy and why it is more "limited" (Bendazzoli, 2018: 2) than corpus-based translation studies (CBTS). Section 1.3 then gives an overview of the different types of existing interpreting corpora, and their characteristics. The different recent research orientations in CIS are explained and illustrated in Section 1.4. Section 1.5 then concludes this chapter with interim conclusions and with the future prospects of CIS.

1.1 Early history

Since the 1980s, corpus linguistics has developed at an accelerated speed. Indeed, while the compilation and exploitation of English language corpora still dominate corpus linguistic research, corpora of other languages have also become available and contributed to the diversity of corpus-based language studies.

Applying corpus linguistic techniques and methods to translation studies was first discussed by Mona Baker who predicted that "the availability of large corpora of both original and translated text, together with the development of a corpus-driven methodology will enable scholars to uncover the nature of translated texts as a mediated communicative event" (Baker, 1993: 243). She then suggested a specific research agenda that involved the design and analysis of parallel, bi/multilingual and above all monolingual comparable corpora. Her first findings were gathered among other contributions in the 1998 special issue of the Translation Studies journal *Meta* edited by Laviosa. In addition to the different types of corpora that could be compiled such as suggested by Baker, the special issue highlighted the potential benefits of using corpora in translator training (Zanettin, 1998) and the advantages and disadvantages of parallel corpora (Malmkjaer, 1998), among many other things. During its first computer-aided twenty years, corpus-based

translation studies (CBTS) was mainly focused on product rather than process and explored linguistic characteristics of translations as texts, such as the proportion of lexical to grammatical words and high- to low- frequency words, word repetition, and type-token ratios.

In 1998, it was clear that corpus-based studies on written language were still far more advanced than studies on spoken language mostly due to the time-consuming nature of data collection and transcription and to paralinguistic dimensions such as rhythm and intonation used by the speaker (Shlesinger, 1998). Today, there is still a gap between CBTS and corpus-based interpreting studies (CIS), both in terms of corpus size and availability as well as in terms of the number of studies and pedagogical applications (Laviosa, 2002; Zanettin et al., 2003 Kruger, 2004; Aston et al., 2004 in Bendazzoli & Sandrelli, 2009). The challenges that I have mentioned partly explain the gap existing between the two disciplines.

Before talking about CBTS in more detail, it is worth defining what a corpus is. The *Oxford Concise English Dictionary* defines a corpus as a “large collection of written or spoken texts”. More specifically, the term *corpus* is associated with at least four characteristics: electronic form, size, representativeness and open-endedness (Fernandes, 2006). With the advent of the computer, a corpus nearly always implies a collection of texts which are represented under an electronic form and can be read and analysed automatically or semi-automatically rather than manually (Baker, 1995). The size of the corpus also matters but historically, corpus-based studies have always relied on huge amounts of data and therefore, the term *corpus* has frequently been associated with big quantities of data extracted from large collection of texts. Nevertheless, in CBTS, a *corpus* can also represent what is known to be “small-scale corpora” (Pearson, 1998). As for representativeness, the choice of texts in a representative corpus depends not only on the size but also of what the corpus intends to represent (Halverson, 1998; Kennedy, 1998 in Fernandes, 2006). Finally, a corpus in CBTS is characterized by its open-endedness which refers to the flexibility that the corpus should have in

order to allow researchers to answer specific research questions. Researchers can then use the texts of this open-ended corpus for different types of comparisons and studies (Olohan, 2004: 48). An accurate way to define a corpus in CBTS is to refer to that corpus as "an open-ended body of machine-readable full texts analysable automatically or semi-automatically and sampled in a principled way in order to be maximally representative of the translation phenomenon under examination" (cf. Baker, 1995 in Fernandes, 2006: 89).

There are different types of corpora being used in the descriptive and applied branches of translation studies. According to Baker's (1995) terminology, there are three main types of corpora: monolingual comparable corpora, parallel corpora and multilingual comparable corpora. In the following paragraphs, I will zoom in on monolingual comparable and parallel corpora. On the one hand, a monolingual comparable corpus can be defined as a corpus that contains components that are collected using the same principles, e.g. the same genres in the same domains written at the same period and in the same language. On the other hand, a parallel corpus can be defined as a corpus containing source texts and their translations. They can be bilingual or multilingual. They can also be uni-directional (e.g. from English into French or from French into English only), bi-directional (e.g. English source texts with their French translations and French source texts with their English translations) or even multi-directional (McEnery & Xiao, 2006).

In the last thirty years, corpus-based translation studies have taken advantage of the fast-growing field of corpus linguistics (CL). As for corpus-based interpreting studies (CIS), it is a quite new branch of Interpreting Studies that has started to emerge in recent years, building on what has already been done in CBTS. Some researchers studying interpreting have been examining performance data but have been facing challenges specific to the field. In fact, the conditions of reception and the production of the translation make the collection and the transcription of oral data difficult for the researcher.

It is Miriam Shlesinger who, in 1998, was the first to publish a paper on the possibilities to extend the corpus-based approach to interpreting and to use already available monolingual corpora to test hypotheses. Shlesinger (2008) refers to CIS "as an off-shoot of corpus-based translation studies". Shlesinger's 2008 seminal paper can be considered as the cornerstone of CIS. It is worth mentioning that Setton (2011) lists all the interpreting studies based on authentic corpora and at least five studies were carried out before the publication of Shlesinger's paper. Indeed, Setton mentions Oléran and Napon's work (1965) as the first corpus-based research endeavour and the author also mentions Déjean Le Feal's (1978), Chernov's (1979), Lederer's (1981) works as interpreting studies based on authentic corpora published before Shlesinger's seminal paper. Yet, it is important to add that these studies were manual, as corpus linguistic tools were not in use at the time.

1.2 CIS: a cottage industry

Today, CIS is still considered "a cottage industry" (Setton, 2011: 34), which means that CIS is more "limited" (Bendazzoli, 2018: 2) compared to CTS research, especially due to small corpus size. However, corpus-based research in the field has been carried out quite a lot over the last twenty years. The volume of authentic conference interpretation recorded and/or transcribed and analysed since the beginnings of interpreting research is rather modest (Pöchhacker, 1994; Kalina, 1998; Monacelli, 2005; Bendazzoli, 2010). Yet, many small samples have been collected and recorded in experimental conditions and many researchers have studied student productions which are much easier to obtain than professional interpretations (Ding et al., 2005).

Recent developments, such as new software, access to larger and quality corpora, and new techniques for transcription and analysis enhance the data but the interpreting studies community still struggles to analyse and understand the practice of interpreting, because they only have primitive theoretical baggage at their disposal, which mainly consists of categories, assumptions and models which

were inherited from descriptive linguistics and cognitive psychology (Setton 2011: 33). Furthermore, interpreting research is mostly concerned with cognitive and psycholinguistic processes and it is really difficult to model the process if factors like live context or features of live speech, such as prosody, are not taken into consideration.

Yet, research methods and doctrine have been a source of debate since the beginnings. On the one hand, the Paris school (Seleskovitch, 1975; Lederer, 1981; Donovan, 1990; Laplace, 1994 in Setton, 2011) insists on the fact that the only valid basis for research is the experience and observations of professionals. From the 1980s onwards, that doctrine was harshly challenged on methodological grounds and controlled experimental studies took the advantage over corpus-intensive research. In the 1990s, CIS had a more sociological paradigm pursued by researchers based in Central Europe and Scandinavia (Pöchhacker, 1994; Kalina, 1998). As CIS was still focused on more controlled experimental studies, it was hard to obtain substantial authentic data. But since the 21st century and thanks to the availability of large multilingual datasets from the European Parliament and to the recent developments mentioned above, CIS is entering a new era.

1.3 Corpus compilation

In this section, I will give an overview of the different types of existing interpreting corpora, their size, the languages used, the interpreting mode and the accessibility of data. It is clear that the compilation of corpora in interpreting studies has changed over time. We can distinguish between three broad categories of corpora (Bendazzoli & Sandrelli, 2009) - manual corpora, early machine-readable corpora and fully machine-readable corpora. Each category will be defined and examples of such corpora will be given. Until not long ago, most of the studies based on corpus data in interpreting were based on traditional or 'manual' analyses because they did not take advantage of computational linguistic or corpus linguistic methods. These studies were also based on small samples,

which were not available in electronic form. In other words, they were not suitable for the automatic data extraction. In his paper, Setton (2011) lists numerous manual projects and it is very likely that the projects dating before 2000 were not machine-readable. Setton focused on studies based on authentic corpora, i.e. empirical data from real life interpreting assignments, so it means that anecdotes or experiments were not taken into consideration. Oléran and Napon (1965), Déjean le Feal (1978), Lederer (1981), Donovan (1994) and Pöchhacker (1994) are perfect examples of manual corpora. Then, we can make a distinction between early and fully machine-readable corpora with the former being not available to the scientific community, contrary to the latter. Here are three examples of early machine-readable corpora:

1) **Fumagalli** (1999-2000) compiled a parallel corpus of 18 English source speeches on international current affairs and corresponding Italian target speeches interpreted by trainees, and a comparable corpus of 15 Italian speeches. Her corpus intended to verify if the main trends of *translationese* (see Baker, 1996) could be identified in interpreted speech. The corpus is not openly available to the scientific community.

2) **Vourikoski** (2004) compiled a corpus of 122 speeches in four different languages recorder at the European Parliament (EP). The transcripts of these speeches and their target versions were available in electronic form, but they would probably need further processing if they were to be analysed with corpus linguistic computer programs.

3) **Straniero Sergio** (2007) recorded a number of interpreter-mediated events on Italian TV in order to study talk-show interpreting.

Summing up, the first attempts to compile corpora in CIS were first manual. Sample data and transcripts could not be analysed with corpus linguistic methods. Then, more steps were taken towards fully-fledged machine-readable corpora with

easier access to recordings. Nevertheless, the general access to these electronic corpora was limited and most projects remained isolated.

A number of more recent corpora are available to the research community and are machine-readable. These machine-readable corpora can be tagged thanks to different software programs, e.g. Treetagger and CLAWS for Part-Of-Speech (POS) tagging, i.e. the classification of words into their parts of speech (see Bendazzoli & Sandrelli, 2006; Dayter 2016). In this section, two fully machine-readable CIS projects are presented, namely **EPIC** and **DIRSI**, even though it is also worth mentioning **FOOTIE**, among others, which is a much more restricted corpus in terms of topics discussed. All texts included in **FOOTIE** indeed come from one type of communicative event, namely the press conferences that took place before and after each game played by Italy's national team during the 2008 European football championships (Bendazzoli & Sandrelli, 2009).

In January 2004, a CIS research group was set up at the University of Bologna at Forlì. Their aim was to study conference interpreters' strategies across different language pairs and directions. To do so, they collected **EPIC** (*European Parliament Interpreting Corpus*), an electronic parallel corpus of source and target speeches in Italian, English and Spanish (Monti et al., 2005; Bendazzoli & Sandrelli, 2005). The plenary sittings of the European Parliament were chosen as source material in this corpus because they show a high level of homogeneity, in that all speeches are produced in the same formal setting (Marzocchi & Zucchetto, 1997). EPIC allows both parallel and comparable analyses and contains about 180,000 words. It is available online for the whole interpreting community to help to share our knowledge of interpreting and even enhance its teaching. As for **DIRSI**, i.e. *Directionality in Simultaneous Interpreting* (Bendazzoli & Sandrelli, 2009), it includes interpreters' output into both their native language and their foreign working language. This corpus was compiled thanks to audio recordings from international conferences about health-related subjects held in Italy between 2005 and 2008 and includes recordings from different sessions (i.e. opening statements, presentations, and closing sessions). The language pair is English

and Italian and five professional interpreters have accepted to contribute data. Debates were excluded from the corpus due to their degree of interactivity. The creation of this corpus has been possible thanks to the experience previously gained with the EPIC project.

It is important to add that interpreting corpora created ad hoc by individual researchers for manual analysis are still used today, and complement the realm of CIS. The notion of corpus in interpreting studies has clearly initially been linked to empirical research based on authentic data (i.e. from real-life interpreting assignments) and because of the difficulty to compile electronic corpora that notion still applies to data sets that continue to be analysed manually and not electronically.

In the following paragraphs, I will describe the main characteristics of interpreting corpora (i.e. interpreting mode and setting, corpus size, languages and data accessibility) that have already been compiled recently, and that are currently being compiled.

The first CIS projects focused on professional simultaneous interpreting performed in conference settings. Two specific sources of data have been predominant, namely TV broadcasting and the European Parliament (EP). At the EP, for instance, source speeches are interpreted simultaneously into as many as 23 languages (sometimes through relais interpreting) and these speeches can be used for research purposes (Bendazzoli, 2010). Other fields have yet been explored such as festivals, medical conferences, and football press conferences. Simultaneous interpreting is nevertheless not the only interpreting mode that can be analysed. Asian research centres focus more on consecutive interpreting because of data accessibility, their data source being televised press conferences of Chinese political representatives (Wang, 2015). More recent projects also focus on short consecutive interpreting in community settings or on dialogue interpreting. It is also important to note that efforts are being made to develop sign-language

corpora despite the difficulty to collect data due to anonymity issues (see Metzger & Roy, 2011).

If they are compared to the spoken part of the British National Corpus (10 million words), interpreting corpora are quite small (Dembry & Love, 2015). Projects based on EP data still have to reach the size of general reference corpora, and it might just be a matter of time and of labour force. However, it is hoped that CIS projects will develop in other international organizations than the EP to diversify interpretation settings. It is already considered in some organisations such as the European Commission (Spinollo, 2018; Scardulla, 2016). Even though current projects carried out in Asia are expected to generate pretty large resources, compiling very large corpora in the near future is not likely to happen.

In terms of languages used, we can see a wide range of language combinations, which confirms one of the "special challenges" (Setton 2011: 68) of CIS - multilingualism. English is represented in many studies, but it is really encouraging to see that non-European languages such as Hebrew, Japanese or Chinese are represented as well.

The last feature of interpreting corpora I would like to focus on is data accessibility, which has always been an issue in CIS. In the oldest studies mentioned in Setton's (2011) overview (see Appendix 1), transcripts were rarely made available and sound files were recorded on tape and not in digital form. Transcript files were thus hard to access (Diriker, 2004) and the analysis had to be carried out manually. In the same period, some studies were however based on machine-readable corpora (Cencini, 2002; Fumagalli, 2000) but the transcripts were "not available for outside use" (Setton 2011: 40). In some other cases, transcripts are available on CDs (e.g. Vuorikoski, 2004; Monacelli, 2009) or on the web as it is the case for the two corpora I mentioned earlier, EPIC and DIRSI. In the future, data accessibility should be facilitated at least among the research community.

1.4 Recent research orientations in CIS

In interpreting, comparisons can be made between source and target speeches, and between written texts and interpreted speeches in the same language. These comparisons can also focus on specific language pairs. The analysis can then be either parallel or comparable (Laviosa, 2012). It is also possible to study interpreters' output as a kind of discourse to point out features that make it different from other types of discourse (e.g. speakers working under different conditions and in different settings, delivery rate, etc.).

This section focuses on the most recent research orientations in CIS, such as interpretese and sociological/ideological aspects.

Features of the product: 'interpretese'

Large-scale CBTS analyses of features such as collocations, part-of-speech distributions and lexical richness in monolingual comparable texts have uncovered traces of simplification in translated text - a universal trait of 'translationese' (Baker, 1995; Laviosa 1996, 2000). As for 'interpretese', in other words, interpreters' speech (Fumagalli, 2000), it has been found that interpreting is closer to original speech than to written translation, linguistically-speaking (Shlesinger, 2008; Shlesinger and Ordan, 2012; Kajzer-Wietrzny, 2018). Kajzer-Wietrnzy investigated the optional complementizer *that* (2012, 2018) and compared the frequency of *that* used in native English speeches and in simultaneous interpretations into English, and used in native English speeches and in non-native English speeches. In both studies, the optional complementizer *that* is less frequent in English native speeches, which shows a higher degree of explicitness in interpretations and in non-native speeches.

Factors in performance and quality

Different variables in interpreting performance have already been explored such as topic, speed, disfluencies patterns (see Sandrelli et al., 2007; Bakti, 2009; Collard & Defrancq, 2019a; Plevoets & Defrancq, 2018) or accents.

In terms of topic, interpreters in EPIC are, for example, much more fluent when interpreting impromptu speeches than political discourses, in all language combinations and directions (Sandrelli et al., 2007). As for disfluencies, Collard & Defrancq (2019a) investigate different types of disfluencies (filled and silent pauses, false starts, lengthening). They show that female interpreters produce less disfluencies than male interpreters. Bakti's paper (2009) states that restarts and grammatical errors are the most frequent disfluencies in the examined corpus (of simultaneous interpreters working from English into Hungarian). Grammatical errors might be caused by a cognitive overload and a higher EVS. Finally, Plevoets & Defrancq (2018) focus on the disfluency *uh(m)* and on four informational load indicators, namely lexical density, delivery rate, numbers and formulaicity. The results indicate that the interpreters' cognitive load decreases when ST are more formulaic (less *uh(m)*), but increases with ST lexical density (more *uh(m)*). It is important to mention that these are only examples of what has been done in the study of interpreting performance.

Strategies, SI technique and process modelling

Using time-aligned authentic corpora is essential to understand and to model the cognitive processes and strategies in simultaneous interpreting. Until today it is the machine-interpreting research group at Nagoya (Japan) that has published the most intensive corpus-fed research on interpreting strategies and processes. They published a classification of Japanese-English SI strategies (Tohyama et al., 2006a), which are the control of the utterance speed, the control of delay time, the control of the length of the pause, and the good use of sight translation pattern. These four categories were used as index to categorize the CIAIR (*Center for Integrated Acoustic Information Center*) simultaneous interpretation corpus into appropriate order of difficulty. They also detailed measurements of lag (Ohno et al., 2008) and showed that Japanese-English interpretation has a larger delay than English-Japanese interpretation and that numerals are interpreted quickly regardless of the language pairs. Finally, they conducted analyses of human vs. machine segmentation for SI and stated that human segmentation was more difficult to implement when there were discourse markers (Ding et al., 2005).

Social and ideological aspects

Some CIS studies focus on intertextuality, norms or interpreter-participant interaction (Diriker, 2001). In terms of norms, Diriker lays the emphasis on objectivity and faithfulness. When the interpreter filters out 'unessential information' (mainly paralinguistic elements), he/she makes a choice, the choice to omit. That means that the interpreter is a much more active participant than usually accepted. CIS studies also investigate the use of linguistic and discourse indicators such as lexical density, repetition, metaphor or politeness (Monacelli, 2009). As for politeness, Monacelli's findings (2009) reveal that interpreters tend to have a self-regulatory behavior, such as "distancing, de-personalization and the mitigation of illocutionary force" (2009: 158). They are then in a detached position to the source text and the target text. Finally, some CIS studies focus on the expression of ideology or cultural identity (Beaton, 2007). Beaton's findings suggest that simultaneous interpreters strengthen EU institutional hegemony, especially due to their extensive use of conceptual metaphor strings.

Pedagogical applications

Corpora can be used for training and educational purposes and we can indeed see that there is now an increasingly established tradition in foreign language learning (Braun, 2006; Flowerdew, 2012; Gavioli, 2005; Kohn, 2012 in Bendazzoli, 2018). As an example, EPIC, which is an organized, indexed sample of professional interpretation, is already used for training purposes (even though it must be acknowledged that the speed and the fact that speeches are read out limit the use of the corpus with advanced students) (Bendazzoli, 2010). As for CIAIR, it is used to demonstrate professional strategies and patterns of lag, chunking or self-correction to interpreter trainees through a pedagogical tool (Tohyama & Matsubara, 2006b). However, the use of corpus-based or corpus-driven methods in interpreter training and education is still in its infancy. First of all, having ready-made interpreting corpora at their disposal can help students to obtain data for their MA theses or to carry out small-scale studies which can lead to greater self-awareness of some processes they have encountered in their curriculum (Dal

Fovo, 2011; Sandrelli, 2010). Corpus concordances can also be good materials to increase trainees' awareness of some phraseological units and other language-specific features that may be an obstacle for the trainee (Aston, 2016; Sandrelli, 2010, Bale, 2013). Finally, learner corpora which include interpreting students (Niemants, 2013; Leung & Yip, 2013 in Bendazzoli, 2018) can also be a relevant resource to know the difficulties and challenges faced during the acquisition of interpreting skills and to promote self-assessment among learners. Some repositories of pedagogical material can even be set up and can then be used by trainees so that they are exposed to a wide range of different source and target speeches.

1.5 Interim conclusion and future prospects

Nearly twenty years after the application of CL to TS, CIS is still at a less advanced stage of development than CBTS. It is clear that one of the reasons for this is that CIS scholars have always had to compile their own corpora in view of the fact that they could not rely on ready-made corpora (McEnery et al. 2006: 71). If we look at more recent studies, we see that CIS scholars have continued to build their own corpora, but they tend not to keep them for themselves anymore. It is also important to add that simultaneous interpreting is not the only interpreting mode that is investigated anymore since studies on classic consecutive and short consecutive (dialogue) interpreting have been carried out to date as well. CIS projects do not only study European languages anymore, so we can hope to see more and more projects based on the investigation of non-European languages. It is also important to highlight that data collection and data processing have been made easier thanks to the opportunities offered by technology and also to the collaboration between researchers.

Now the challenge lies in increasing trainees' awareness and finding possible applications to interpreter education and professional practice. Some examples of corpus used in the interpreting curriculum already offer stimulating ways to increase trainees' awareness of collocations, phraseology, language-specific

features and terminological preparation. Small-scale and large-scale interpreting corpora, machine-readable or not, show a great potential as they do not only serve the purpose of one project, but allow the development of interpreter training and professional practice. In the near future, we will be witnessing the creation of corpora linked to different types of interpreting as a result of greater collaboration in data collection, transcription, annotation and sharing.

2 Teaching simultaneous interpreting

In this chapter, the focus is on teaching simultaneous interpreting. After a brief introduction in Section 2.1, Section 2.2 sheds light on the early history of SI training, from the beginnings until 2000. Section 2.3 then explains what Gile's Effort Model is and why it is important to keep in mind as an interpreter trainee. In Section 2.4, seven different difficulties (numbers, proper names, complex noun phrases, singe-word terms, culture-specific items, idioms, and phrasal verbs) of SI are explained and illustrated, with a special focus on English as a source language.

2.1 Introduction

Interpreting is a branch of the larger field of translation. Translation refers to the result of a written activity while interpreting is an oral reproduction of a message (Stoica, 2008). In the field on interpretation, there are three main subcategories, namely consecutive interpreting with note-taking, simultaneous interpreting of speeches as delivered at international events, and dialogue interpreting in community settings. The two first subtypes differ in various ways. First, consecutive interpreting does not require the interpreter to reproduce immediately the message as in simultaneous interpreting, where there is no time for structuring the discourse. Another distinction can be made in the understanding phase. In consecutive interpreting, it can take the form of notes whereas for the simultaneous interpreting, the processing of information must be very immediate, making the task more cognitively challenging. In both cases, the interpreter needs a good short- and long-term memory (ibid.). In this chapter, I will only focus on simultaneous interpreting, and more precisely on interpreter training, the difficulties faced by students and the strategies than can be used so that the source message is successfully conveyed.

2.2 Early history of simultaneous interpreting training

Since the first writings on the subject (Herbert, 1952; Paneth, 1957) more than fifty years ago, the training of simultaneous interpreting has undergone considerable development and institutionalization. Mackintosh (1995: 122-128) suggested that there is now a widely accepted training model which implies, for example, that trainers should be "practicing conference interpreters, preferably AIIC members" and "use material taken from actual conferences as teaching material". In line with Mackintosh's (1995) "Training paradigm", it is also suggested that "good training programs offer students sufficient exposure to a variety of speakers" (Moser-Mercer, 1994: 15). As mentioned in Pöchhacker's (1999) paper, Kurz (1989, 1990) and Cenková (1994) recommend using different types of video recordings such as live, audio- or videotaped speeches, whereas Schweda Nicholson (1993) and Schjoldager (1996) describe correction techniques and suggest feedback sessions.

In general, contributions to the didactic literature on interpreter training often take the form of report by teachers willing to share their particular approach (which can be descriptive or prescriptive). The prescriptive approach is used in textbooks on interpreter training (e.g. Seleskovitch & Lederer 1989, 1995; Gile, 1995). Interpreter training is experiencing a growth, especially in China, as more didactic issues are being investigated through systematic empirical research.

2.3 Gile's effort model of simultaneous interpretation

In the seventies, some models on information-processing paradigm were developed to account for the mental operations of SI (Gerver, 1976, Moser, 1978). More recently, models on cognitive science, such as Setton's (1997), have also been developed. All of these are valuable as they take into consideration all the relevant developments in cognitive psychology, neurolinguistics, and linguistics (Gile, 1999). However, they have not been subjected to much testing over the last two decades, probably due to a lack of resources and due to the complexity of the mental operations involved.

As said in Section 2.2, Daniel Gile has a prescriptive approach of interpreter training and proposed his Effort Model to help interpreters understand the "difficulties [of interpreting] and select appropriate strategies and tactics" (Gile, 1992: 191). The Effort Model is based on the idea that the brain has a limited capacity and that the difficulties in interpretation are due to time constraints and the need to divide attention between different operations (Gile 1995: 91). This model is based on three non-automatic interpreting Efforts (see Gile 1999: 154, Setton 1999: 35 for a review):

1) **Listening and Analysis Effort**: it includes all the mental operations between the perception of discourse and the moment at which the interpreter decides to say or not say what he/she heard.
2) **Production Effort**: it includes all the mental operations between the moment at which the interpreter decides to convey the meaning and the moment at which he/she actually formulates it in the target speech.
3) **Working Memory (WM) Effort/ Short-Term Memory (STM) Effort**: it includes all the memory operations from the time a segment is heard to when it is reformulated in the target speech or disappears from the memory.

The sum of the three Efforts represents the total capacity required, which cannot exceed the available capacity. The Coordination Effort, which is the management between the three Efforts mentioned above, causszes further energy consumption for the interpreter. If one of the three Efforts needs more resources at one moment, a part of the capacity is diverted from the others. However, there is a cognitive overload whenever the capacity is insufficient, either in terms of the individual efforts or for the cognitive system as a whole.

2.4 Difficulties of simultaneous interpreting, with a special focus on English as a source language

In this section, I will focus on the difficulties of simultaneous interpreting. As mentioned in Gile's Effort Model, each Effort demands resources to work properly.

There are different kinds of operations that could make processing demands exceed the available capacity, leading to deterioration in the content and/or the form of the interpretation. As interpreters generally work at the limits of their cognitive capacities (Gile, 1999), these information overflows are common. In this study, I decided to work on different items of the English language that, in my opinion, may be difficult for interpreters working from English into French. These items can be found in different languages even though some of them are specific to English, and are the following: numbers, proper names, complex noun phrases, culture-specific items, idioms, phrasal verbs and single-word terms. Each subsection will provide an analysis that describes the potential difficulties of this type of items and suggest some strategies to cope with those difficulties.

2.4.1 Numbers

In everyday practice, conference interpreters admit having some difficulties with numbers, with error rates reaching 40% for professional interpreters and 40% to 70% for trainees (Korpal, 2016; Korpal & Stachowiak, 2018; Timarová, 2012 in Collard & Defrancq, 2018). It can be surprising that numbers are such a source of problems as their main characteristic is univocity of meaning (Alessandrini, 1990). Their interpretation should be quite easy. In most modern languages, the system developed to create numbers makes it possible to create all numbers using a small set of basic words and fixed syntactic rules, each providing a single piece of information on the number. For example, in English, the main constituents are called base digits (0 to 9 in the decimal system), which can be used to creates two lexical classes: teens (*eleven, twelve... nineteen*) and tens (t*wenty, thirty... ninety*). There is a fourth set of items, called multipliers, (*hundred, thousand*) used to mark the magnitude of the base digits. This system very often requires a lot of items and words to create a large number in the target language and therefore the problem encountered by interpreters can be a memory problem due to the need to remember many words for the expression of a single concept (Mazza, 2001). Numbers are thus characterized by high information density and also low

predictability, which requires the interpret to change strategies with respect to listening, memory and production (De Smet, Vandierendonck & Defrancq, 2018).

Even though interpreters broadly agree that numbers are particularly difficult to interpret, research on the topic is rather limited (Mead 2015: 287). If we look at studies on numbers, Gile mentions them several times in his works (1995, 1999) and highlights the difficulty of numbers in SI because of three different characteristics, which are mentioned by many other researchers:

1) **Low predictability** (Braun and Clarici, 1996; Mazza, 2001; Pinochi, 2009, Mead, 2015): it increases the Listening and WM Efforts because the anticipation is rarely possible (Jones, 2002; Pinochi, 2009). The number expressed can only be understood once the speaker utters it, and anticipation is therefore almost impossible.
2) **Low redundancy**: it also increases the Listening and WM Efforts because it is really important not to miss or forget them as they cannot be inferred from any element in the text (Gile, 1995: 108).
3) **High informative content** (Alessandrini, 1990): sections with numbers increase processing capacity demands for all Efforts as the interpreter must process, remember and translate more information.

These characteristics can explain the frequency of mistakes with numbers in SI, especially large numbers which are expressed using several words and can also explain why interpreters cannot use strategies such as paraphrasing or reformulation (Jones, 2002; Pinochi, 2009) as every component of a number is a unit that represents one meaning only.

Some strategies can, however, be applied to reduce the number of errors linked to numbers. One possible solution to deal with numbers is the reduction of *decalage,* i.e. a shorter ear-voice span (EVS)[1], which allows shorter retention in STM but faster restitution of the input and also allows the interpreter to stick as close as

[1] The ear-voice span is the time lag between a word uttered by the source speaker and the corresponding word uttered by the interpreter (Defrancq & Collard, 2018)

possible to the speaker. It is suggested that professional interpreters perform better than students when dealing with numbers because they keep a shorter EVS (Timarová et al., 2014). Gile (2008) finds this strategy logical because a short EVS means that the interpreter saves working memory capacity. Kalina (1992: 254) suggests another strategy which is *approximation*. It can be a useful strategy to deal with particularly complex segments. She thinks this strategy can be used to provide partial information until the interpreter finds a more accurate translation, but this might be counterproductive for numbers as strategies to gain time increase the WM Effort.

Finally, since the main problem with numbers seem to be insufficient WM, writing the numbers down while interpreting might also be a good strategy (Mead, 2015; Setton, 1999). Despite the fact that note-taking can represent another Effort which will want its share of the total attentional capacity, it might relieve pressure on WM. A few studies (Crevatin, 1991; Braun & Clarici, 1996) have focused on the management of numerals in SI and the usefulness of note-taking. They showed that notes were useful for numbers during the interpretation into the A-language even though the performance of professional interpreters and students was relatively poor. Furthermore, the language pair was not the same as in this dissertation, namely Italian-English or Italian-German for the two studies mentioned. However, Mazza carried out another study (2001), which showed that taking notes when dealing with large numbers (four digits or more) did not help the interpreter because large numbers are often too dense to be rightly noted.

2.4.2 Proper names

Seleskovitch (1962) stated that names require linguistic “transcoding” rather than interpreting. Gile highlights that proper names are potential problems for interpreters, particularly if the interpreter is not familiar with a specific proper name or its pronunciation in the target language (Gile, 1995). Proper names may also increase the efforts of the interpreter, and thus require some “coping tactics”. Proper names can be phonetically unfamiliar to the interpreter (Hanaoka, 2002).

In that case, interpreters have to store the names in their short-term memory and recode them in their target language. However, the lack of schemata (a schematic view of the knowledge of proper names) may affect their capacity to retain the name in their short-term memory (Rumelhart, 1980). Interpreters may also run the risk of mispronouncing the name if they do not know how the word is spelled beforehand (Quini, 1993) and may sometimes not realize that there is a conventional translation for a specific name in their target language.

Hanaoka (2002) proposes a distinction between two types of strategies in order to render proper names more easily: **decoding strategies** and **encoding strategies**. The first strategies consist of being able to recognize a proper name and gaining knowledge about the referent. There are also two types of decoding strategies. First, it is possible to put proactive strategies into place if the interpreter can anticipate the topic of the text. An interpreter can for instance search for specific proper names that he/she can encounter if given the topic. Conference interpreters are usually informed of the topic of the conference in advance and can therefore do research on that topic. Then, reactive strategies are employed when the interpreter has encountered an unfamiliar name, and they include guessing from the context or the use of an online dictionary. However, in simultaneous interpreting, the resources may be significantly limited due to time constraints.

As for encoding strategies, three factors need to be considered (Hanaoka, 2002):

1) **The culture-specificity of the name**: the interpreter has to decide how familiar the name may be to the listener, so their assessment of the listener's knowledge state is a crucial factor in their choice of strategy.
2) **Relationship to macro/micro-structure**: if the referent of a name is the topic of a speech, then it is part of the macro-structure of the speech and is essential. However, a name considered less important may be more susceptible to simplification strategies.
3) **Topical knowledge/figurative knowledge**: when an interpreter faces an unfamiliar proper name due to a lack of cultural knowledge, the addition or explication of extratextual factual information is possible. In the case of a

figurative use of a name, the explicitation of the figurative meaning is required.

2.4.3 Complex noun phrases

Complex noun phrases (NPs) are common in English but exist in every language in the world and generally consist of a head, normally a noun, and of elements that determine and optionally modify the head or complement another element of the phrase. Here are some examples:

(1) *expensive* ***houses*** (modifier + head)
(2) ***houses*** *downtown* (head + modifier)
(3) *expensive* ***houses*** *downtown* (modifier + noun + modifier)

Complex noun phrases can be potentially difficult for interpreters in simultaneous interpreting (Barbafina, 2003; Ghiselli, 2018) as a series of words that go together can overload the interpreter's working memory. Ghiselli's study (2018), based on professional conference interpreting data, shows that the most common interpreting strategies to cope with complex NPs are complete and correct translations and generalisations.

Therefore, the translation of complex noun phrases is challenging and special attention in training and practice of simultaneous interpreting is recommended. For example, the **control of Ear-Voice-Span** (EVS) is a crucial issue (Ghiselli, 2018). Van Dam (1989) describes a set of beginner exercises which include distance exercises to avoid tailgating the speaker and keeping the optimum start-up distance, which corresponds to approximately one meaning unit behind the speaker at the beginning of each new sentence. EVS has to vary during the task and a balance should be found between understanding the message before speaking and not overloading working memory. Another strategy that might be useful in the case of complex noun phrases is **compression** or **condensing** (Bartlomiejczyk, 2006; Kader & Seubert, 2015). Compression means summarizing a longer fragment with a shorter phrase, which is supposed to convey the same meaning, but expressed in a more concise and general way.

2.4.4 Single-word terms

In my opinion, and as an interpreter trainee, single-word terms are difficult to handle as they are short and refer to a specific concept. A term is a word closely linked to a specialized conceptual content, which relates to a specific area or field or discipline (Khan, 2016). The dictionary of Linguistics and Phonetics defines a term as "a unit of expression which has universal intuitive recognition by native speakers, in both spoken and written language". Ananiadou (1994: 1034) says that a term may be a single word or a multiword unit, but in this study, I will only focus on single-word terms. Here are some examples:

(4) *measles* > *rougeole*
(5) *nitrogen* > *azote*

There is not much literature about strategies when facing single-word terms in the interpreting field. Khan (2016) suggests that teaching terminology during the university cursus will help translators and interpreters. According to Khan, planning terminology classes will help learners to coin new terms as they can be in critical situations where equivalents may not be found in dictionaries.

2.4.5 Culture-specific items

Culture-bound terms, or culture-specific terms (Harvey, 2000), refer to concepts, institutions and items which are specific to the source language culture and are classified into five categories (Newmark, 1988)[2]:

1) **Ecology**: *selva* (tropical rain forest), *plateau, savanna, paddy field*
2) **Material culture**: *zabaglione* (food), *anorak* (clothes), *bourgade* (houses and towns), *cabriolet* (transport)
3) **Social culture**: *sithar, reggae, rock* (work and leisure)
4) **Organizations/customs/activities/procedures/concepts**: *karma, temple* (religious)
5) **Gestures and habits**: *cock a snook, spitting*

[2] For more examples, see Simunic (2013: 14).

There is a general consensus that culture-specific items are potentially problematic for interpreters (Amato & Mack, 2011). It is important for interpreters to have basic cultural knowledge of the source language culture as culture-specific terms are common in speeches. They often have no equivalent in the target listener's cultural frame, which makes them difficult to translate into the target language (Simunic, 2013). The interpreter has to choose among different interpreting strategies to deliver the message. The following paragraph presents these different strategies and gives examples for each category, which are from Simunic's (2013) paper (language pair: English to Croatian).

The strategies used in simultaneous interpretation are the following[3]:

1) **Using a term/word/phrase of similar meaning and form**: it consists of using an item which conveys roughly the same meaning as that of the source language item. The item used in the target language consists of equivalent lexical terms. This technique is only occasionally achieved.

(6) *shared space* > *dijeljenje prostora*

2) **Using a term/word/phrase of similar meaning but dissimilar form**: it consists of using an item which has a similar meaning as the source item, but which is made of different lexical items.

(7) *dozen days* > *nekoliko dana* (*tucet* is the literal translation)

3) **Paraphrase**: a paraphrase is an explanation of a unit of language (Newmark, 1988) used to clarify an obscure phrase. This is the most common way to translate terms/words/phrases when an equivalent cannot be found in the target language (Baker, 1992). Interpreters show their ability to cope with an unfamiliar item under pressure.

(8) *white-collar and blue-collar jobs* > *uredski i zanatski posao*

[3] Strategies 1-4 are proposed by Mona Baker (1992, 2006) and strategies 5-9 by Harvey (2000), who adapted his strategies from five procedures identified by Weston (1991: 19-34)

4) **Omission**: an interpreter may choose to omit an item because it has no close match in the target language or because its meaning cannot be easily paraphrased or because it is of no great importance for the listener.

(9) *National Health Service* > omission

5) **Functional equivalence**: this technique consists of using a term/word/phrase in the target language whose function is similar to the item used in the source language. However, authors are divided over the efficiency of this technique which is, on the one hand, considered as misleading (Sarcevic, 1985) and, on the other hand, as the ideal method of translation (Weston, 1991).

(10) *Breton shortcake* > *bretonski keks*

6) **Formal equivalence** (also called 'linguistic equivalence'): this technique is mainly used when translating certain legal, political, sports or any institutions or organizations which exist or existed at some point in the target language culture.

(11) *green paper* > *zelena knjiga*

7) **Transcription/borrowing**: it consists of reproducing or, when necessary, phonologically adapting the original term/word/phrase. This strategy is mostly used if the culture-specific term cannot be interpreted by any other strategy because it does not exist in the same form in the target language culture.

(12) *businessmen* > *biznismeni*

8) **Conventionalization**: this strategy is used to translate proper names which have an established name in the target culture.

(13) *Cold War* > *Hladni rat*

9) **Descriptive translation**: it consists of using generic words rather than culture-bound terms to convey the meaning. This technique is appropriate when formal equivalence is considered not clear enough.

(14) *Globish* > globalni engleski

2.4.6 Idioms

As an interpreter trainee, I find it rather challenging to handle English idioms, such as (Baker, 2006: 65):

(15) *throw caution to the winds*
(16) *storm in a tea cup*
(17) *blow someone to kingdom come*
(18) *like water off a duck's back*

Idioms are "linguistic expressions or lexical items representing objects, concepts or phenomena of material life particular to a given culture" (Dastjerdi, 2011: 879). Gottlieb suggests that "an idiom is difficult to decode correctly for someone who only knows the normal meanings of its constituent elements" (1997: 260). Idioms can include many different cultural aspects such as religious beliefs, culture-specific items and superstitions. Every society perceives the world in a different way and that vision influences the language. As a result, the collocations and idiomatic expressions of the languages diverge from each other in most cases (Shojaei, 2012). This difficulty has not been in IS, but some authors have already carried out studies about the difficulty of translating idioms in TS. Even though it would be more relevant to list strategies for the interpreter, it would be interesting to see if the strategies that should be used by translators can also be used by interpreters, considering the specificities of this form of interlinguistic mediation.

When the translator is faced with an idiom, he/she should investigate which strategy is more appropriate in dealing with the translation of the source idiom into the target language. Baker (1992) suggests four problem-solving strategies to help translators and also gives examples for each strategy:

1) **Using an idiom of similar meaning and form**: it is the least used strategy since languages radically differ in the way they identify a single concept.

However, it is considered as the ideal strategy for translating idioms as an exact equivalent is found in the target language.

(19) *the rain fell on the just and on the unjust* > *la pluie tombait aussi bien sur les justes que sur les injustes*

2) **Using an idiom of similar meaning but dissimilar form**: in this case, the meaning is the same, but the lexical items are different.

(20) The serow, a type of wild mountain goat, is *very much at home* among the rocky outcrops of Sichuan > The serow, a type of wild mountain goat, is *totally at ease* in Sichuan's many rocky levels (back-translation from Chinese)

3) **Translation by paraphrase**: this is the most commonly used strategy when the translator cannot find any equivalents for the source idiom. Presenting more clarifications on the idiom is a wise choice. However, there is a danger of losing the intended effect and the cultural significance (Baker, 1992: 74).

(21) This is rather like *shutting the stable door when the horse has bolted* > c'est peut-être *trop peu trop tard*

4) **Omission**: this strategy is used to completely omit the idiom from the target text when there is no close match between the languages' items or when the translator cannot find any equivalents.

(22) It was bitter, but funny, to see that Professor Smith had doubled his own salary before recommending the offer from Fayed, and added a pre-dated bonus *for good measure* > It was regrettable, even funny, that Professor Smith had been able to double his salary twice before offering his recommendation to accept Fayed's offer, and that he added to this a bonus, the date of which had been previously decided on (back-translation from Arabic)

2.4.7 Phrasal verbs

There are different kind of verbal formulaic expressions such as prepositional verbs, phrasal verbs, phrasal-prepositional verbs and other multiword verb constructions (Biber et al., 1999: 403-427). However, among these, phrasal verbs are the most frequent forms of multiword verbs. They are formed of a lexical verb in combination with an adverbial particle, and here are some examples (Oluchukwu, 2015):

(23) *to look up*
(24) *to run into*
(25) *to back out*
(26) *to call round*
(27) *to egg on*
(28) *to brush up*

They are fairly frequent in the English language, occurring in the *British National Corpus* at a rate of approximately one every 192 words (Gardner and Davies, 2007: 347). One of the challenges of phrasal verbs is that it is hard to find a phrasal verb in the target language that belongs to the same grammatical category to convey the same meaning as in the source text (Oluchukwu, 2015). In this regard, Baldwin & Kim (2010: 267) state that “the lexical and syntactic and semantic idiosyncrasies of multiword expressions are at the root of translation problems”. To date, the issue of rendering phrasal verbs in interpreting has not been discussed in CIS.

2.5 Research questions

In this dissertation, I would like to address two main research questions. Considering the difficulties inherent in the English language described in Section 2.4., my first research question is the following:

> Which source linguistic items are the most challenging for interpreter trainees working from English into French? In tackling this question, I also aim to show how it is possible to determine that a specific item is challenging or difficult for an interpreter.

Operationalizing interpreting quality as a low number of errors (i.e. divergent renditions), I would like to address a second research question:

> Does the quality of the trainees' interpretations improve across time? How can we measure that improvement? Do all trainees improve across time?

3 Data and methodology

This chapter is divided into four sections. Section 3.1 gives information about the interpreter trainees and explains how the corpus used in the present study was compiled. The transcription guidelines used for my data are explained and illustrated in Section 3.2. Section 3.3 gives information about the corpus used, such as source speech topic, source speech length, or the total number of running words in each source and target text. Finally, the construction of the database on which the analyses are based is explained in Section 3.4.

3.1 Corpus collection

In order to analyse the differences between a given source speech in English and its interpretations into French, I had to collect relevant data. That is why I recorded and transcribed the performances of eight master's students of the UCLouvain (Belgium), named Student 1 to Student 8, who were all in the process of getting their master's degree in conference interpreting. I recorded four of their simultaneous interpretations in November and December 2016, across four different sessions of their simultaneous interpreting course. They filled in a consent form (see Appendix 2) before the interpreting tasks and knew they were being recorded. I have not collected metadata related to the students' gender, age, or language proficiency because I knew from the start that these factors would not be investigated in the present case study. However, it is important to say that the students were all native speakers of French. During these four interpreting sessions, their teacher, a non-native speaker of English (NNS), was reading out texts that were transcripts of oral speeches, and the students were alone in the interpreting booths. In other words, they did not have anyone to help them out when they were facing a difficulty. The four speeches were rather long, between 30 minutes and 1 hour, and I therefore decided to study the first twelve to fourteen minutes of each speech, starting at the very beginning and stopping when the lecturer was finishing the idea that was uttered around the minute thirteen. Every

performance was then transcribed according to the EPTIC guidelines. More information about the transcription process and about these guidelines can be found in Section 3.2.

The eight students did not attend all four classes, which is why I have collected and transcribed twenty-two outputs in total. Table 1 shows the number of outputs per student. Student 1 and Student 2 attended the four sessions while Student 8 only took part in one of them. It will be interesting to investigate if there is some kind of improvement across time (see research question 2) for students who attended at least three out of the four classes (Students 1, 2, 4, 6 and 7).

Table 1 – Distribution across speeches and students

	Speech 1 (3rd November)	Speech 2 (10th November)	Speech 3 (17th November)	Speech 4 (8th December)	Total
Student 1	X	X	X	X	4
Student 2	X	X	X	X	4
Student 3		X	X		2
Student 4	X		X	X	3
Student 5			X	X	2
Student 6		X	X	X	3
Student 7	X	X		X	3
Student 8	X				1
Total	5	5	6	6	22

In the present dissertation, the first speech will be referred as IN01, which stands for “input number 1” and the same principle applies to the other source speeches. The first student will be referred as STU01, which stands for “student number 1” and the same principle also applies to the other students.

3.2 Data transcription

In corpus-based interpreting studies, as in any corpus investigation that looks at spoken language, it is necessary to transcribe oral data in order to turn them into an analysable format. However, as Agar (1996: 153) says, "transcription is a chore" and transcribing interpreting data is not easy neither. Niemants (2012: 180) furthermore says that "interpreting research is not very helpful when it comes to determining an adequate transcription system". When it comes to transcribing interpreting data, Niemants talks about three main issues: participants, languages, and timing. Any transcription involving more than one participant needs to the speaker of each utterance, the beginning and the end. The main problem about **participants** is then to segment utterances and link originals with their renditions, knowing that one original segment does not systematically correspond to one rendition segment and vice versa. **Multilingualism** is also a big issue for interpreting data as the transcriber must be able to input language-specific characters and specify the language used. There are also many phonetic and prosodic features to take into consideration: tempo, volume, duration, pitch, intonation, and accent, among others. Finally, **timing** is also a big issue as we have to pay attention to the beginning and the end of all utterances, and also to how they intertwine (Niemants, 2012).

For the transcription of my interpreting data, I followed a set of guidelines, which are described in Table 2. These guidelines are used for the transcription and for the enlargement of EPTIC, and they serve those involved in the process of transcription and coding.

Table 2 – EPTIC guidelines

Feature	Example	Transcription convention
Punctuation		Based on the syntax and the intonation, no comas
Truncated words	propo	propo-
Pauses	filled	ehm

	empty	…
Interrupted speech		…
Numbers	Numbers	Written out in full
	326 000 000	326 millions
	326 000 000 000	326 milliards
	30 %	30 pour cent
	Years	In numbers
	Fractions	Written out in full
Unclear bits		#
Language shift		#
Invented words	endavure	{endavure}
Capital letters		To refer to interinstitutional stylistic guidelines and official websites such as: http://www.cnrtl.fr/definition/
Acronyms		To refer to interinstitutional stylistic guidelines and official websites
Titles		monsieur/madame/mademoiselle

First, all the words that the students started uttering but did not complete (i.e. truncated words) are transcribed with a hyphen.

(29) *Le don- don- le donim- le dénominateur commun est simplement ce cette inflexion cette intensité*

There are then two types of pauses: empty and filled pauses. An empty pause is transcribed using "…" (irrespective of its length) whereas a filled pause is transcribed using "ehm". A **filled pause** is "a period of articulation of non-propositional content and fitting a language specific convention" (Rose, 2017 based on Mahl, 1956; Clark & Fox Tree, 2002). In English, "um" and "uh" are typical filled pauses. Then, the symbol # is used for unclear bits. Non-attested words in the target language are transcribed between curly braces (either totally invented words or English calques that do not exist in French).

(30) {*combinations*}

(31) {*synapsis*}

(32) {*synesthie*}

3.3 Corpus description

This study relies on a multiple *parallel corpus*, defined as "bidirectional translation corpora" (Johansson, 2007: 9-11) or, in other words, a set of texts in language A and their translations in language B. The corpus is named LOCOSSI and is made up of 26 speeches, i.e. 4 source speeches (the English subcorpus) and 22 interpretations (the French subcorpus). It contains 36,154 tokens. The corpus used in the present study is **parallel**, with source texts in English and their interpretations in French. It is thus **bilingual** (two languages represented) and also **unidirectional** (only one interpreting direction, from English into French). It can also be termed a **multiple** parallel corpus in the sense that it contains several interpretations of the same source speeches. Table 3 shows the total number of words per text. The grey row shows the total number of words of the English source texts.

Table 3 – Total number of running words (source and target texts)

	IN01 1,378	IN02 1,343	IN03 1,487	IN04 1,414	**Total IN** 5,622
STU01	1,451	1,460	1,632	1,429	5,972
STU02	1,367	1,322	1,505	1,400	5,594
STU03	/	1,457	1,458	/	2,915
STU04	1,170	/	1,436	1,313	3,919
STU05	/	/	1,641	1,463	3,104
STU06	/	1,086	1,375	1,205	3,666
STU07	1,524	1,334	/	1,216	4,074
STU08	1,267	/	/	/	1,267
Total STU	6,779	6,659	9,047	8,026	30,511

The four source speeches are part of the pedagogical material used by a lecturer in English-to-French interpreting. The students who interpreted these texts into French were all interpreter trainees in their second year of master's degree in conference interpreting.

The first source text (ST) included in LOCOSSI, which was used on November 3rd, 2016, contains 1,378 words (which corresponds to 14:04 minutes). It is the first part of a much longer speech, which contains 3,610 words in total. It is taken from a BBC Radio 4 programme called the Reith Lectures. The text is about synaesthesia, which is a perceptual condition in which the simulation of one sense triggers an automatic, involuntary experience in another sense. For example, synesthetes, people who have synaesthesia, may see sounds, taste words or feel a sensation on their skin when they smell certain scents. The source speech is scientific and accordingly, contains specialized terminology. Here are some examples:

(33) *angular gyrus*
(34) *sickle cell anemia*
(35) *cross-modal synesthetic abstraction*

As ST01 contains specialized terminology, it is a specialized text. A **specialized text** is a production of non-literary pragmatic content designed to be used in a specific field or discipline, such as science, technology, healthcare, business, or tourism. The purpose of a specialized text is mostly informative.

The second ST was used on November 10th, 2016 and contains 1,343 words (duration: 13:37 minutes). As the first ST, it is only a sample of a longer speech made up of 3,464 words. This ST is also from the Reith Lectures BBC programme. The text is about the brain and its complex structure and is a scientific text and can then also be considered as a specialized text (with specialized terminology).

(36) Capgras Syndrome
(37) fusiform gyrus
(38) *Darwinian revolution*

The third ST, used on November 17th, 2016, is made of 1,487 words (duration: 13:45 minutes). The full speech contains 5,471 words and is also based on Reith Lectures. It deals with the evolution of art and how people over the world perceive art. The text contains references to the famous Chola bronze representing goddess Parvati in Southern India. This text does not contain any specialized terminology.

Finally, the last ST was used on December 8th, 2016 and contains 1,414 words (duration: 12:52 minutes). It is the transcript of a TED talk given by Juan Enriquez in November 2016: "What will humans look like in 100 years?" (total length: 3,115 words). The talk is about the evolution of the human body and the ethics behind that evolution. The speaker mentions new kinds of prosthetics, such as hearing aids. ST03 is a specialized text (scientific) and therefore contains specialized terminology.

(39) *gene code*
(40) *hereditary* diseases
(41) *radiation*

The four ST were POS-tagged with the TreeTagger, using Sketch Engine. **Part-of-speech tagging** (POS-tagging) is the action of annotating corpora, or in other words, the process of classifying words into their parts of speech and labelling them accordingly. The **TreeTagger** (Schmid, 1994) is a tool for annotating text with part-of-speech and lemma information. It was developed by Helmut Schmid in the TC project (Stuttgart). Finally, **Sketch Engine** is a corpus query system, which allows the user to view and analyse word sketches (Kilgariff et al., 2004). The POS-tag description used for this subcorpus can be found in the appendices (Appendix 3).

This French subcorpus, made up of 30,511 words, was also POS-tagged, using the TreeTagger for French on the Sketch Engine platform. More information about the French tagset can be found in Appendix 4.

3.4 Database

In this section, I explain how I built the database on which the analyses are based. I will first take a look at the source speeches, focusing on the difficulties they contain and their frequency of occurrence. The second part of this section outlines the taxonomy of rendition types I used. Rendition types are illustrated with authentic examples taken from LOCOSSI.

3.4.1 Source speeches: types of difficulties

As mentioned in Section 2.4., this study focuses on different difficulties faced by trainees in SI. These potentially problematic items are numbers, proper names, complex noun phrases, single-word terms, culture-specific items, phrasal verbs and idioms. I manually identified all the occurrences of each type of difficulty in the four source speeches. All these occurrences were then included in a database (an Excel spreadsheet). Table 4 shows examples of each category. As the "complex noun phrases" category was the most populated (see Table 5), I decided to subdivide it into eight subcategories, which represent different syntactic patterns (where A stands for adjective, N for noun, prep for preposition, and Adv for adverb).

Table 4 – Examples of difficulties

Type of difficulty	Examples (taken from LOCOSSI)
Complex noun phrases	
⇒ A + N	*flaky types, greater hyperconnectivity, bulbous shape, basic anatomy, deliberate hyperbole*
⇒ N + prep + N	*photograph <u>of</u> you, universal laws <u>of</u> art, staggering complexity <u>of</u> the structure, three-dimensional representation <u>of</u> your own body*
⇒ N + subclause	*places <u>where</u>, two recurring themes <u>that</u>, Victorian Englishmen <u>who</u>*

⇒ N + N	*ballroom dance, heart pacemaker, body parts, brain map*
⇒ Adv + A + N	*seemingly unrelated concepts, totally hypothetical example, absolutely extraordinary prosthetics*
⇒ A + N + N	*European Renaissance art, shattered glass shape, different sense perceptions*
⇒ N + N + N	*sickle cell anaemia*
⇨ Other (infrequent patterns)	*property of being jagged, risk of annoying some of specialist colleagues, rough idea of what the brain looks like*
Single-word terms	*cosmos, neuron, synaesthesia, amoeba, cortex*
Numbers	*two shapes, one in two hundred people, in the last three hundred years, first*
Proper names	*Aldous Huxley, Lord Reith, Shakespeare, Angela Merkel*
Culture-specific items	*Chola, goddess Parvati, Yankee, soft-boiled egg*
Idioms	*time and again, in short, wouldn't give me a penny for it, ruling with an iron fist*
Phrasal verbs	*step out, filter out, turns out*

Table 5 shows the number of occurrences of each category and gives a first rough estimate of the difficulty profile of the source speeches.

Table 5 – Difficulty profile of source speeches (tokens)

	IN01	*IN02*	*IN03*	*IN04*	***Frequency***
Complex noun phrases	60	77	64	64	265 (72.6%)
A + N	33	24	31	25	113 (30.96%)
N + prep + N	11	27	15	18	71 (19.45%)
N +	7	18	7	14	46 (12.60%)

subclause					
N + N	2	4	6	6	18 (4.93%)
Adv + A + N	2	1	3	1	7 (1.98%)
other	2	3	1	0	6 (1.64%)
A + N + N	2	0	1	0	3 (0.82%)
N + N + N	1	0	0	0	1 (0.27%)
Single-word terms	12	7	3	12	34 (9.31%)
Numbers	6	11	7	7	31 (8.49%)
Proper names	1	4	8	3	16 (4.38%)
Culture-specific items	0	0	6	2	8 (2.19%)
Idioms	3	1	1	1	6 (1.64%)
Phrasal verbs	2	0	1	2	5 (1.37%)
Total	84	100	90	91	365 (100%)

In total, the English source speeches contain 365 occurrences of the difficulties examined here and each of them contains approximately the same number of potentially problematic items (IN01 has the lowest number of items while IN02 has the highest). Table 5 shows that **complex noun phrases** are, by far, the most frequent category, as there are 265 NPs in the source speeches, representing 72.6% of the difficulties faced by interpreter trainees. The most frequent pattern of NPs is “A + N” (adjective + noun), with 113 occurrences, representing 42.64% of all NPs and 30.96% of the total number of difficulties. The second most frequent category is “N + prep + N” (noun + preposition + noun), with 71 occurrences, representing 26.79% of all NPs and 19.45% of the total number of difficulties. Finally, the third most frequent pattern of NPs is “N + subclause” (noun + subclause), with 46 occurrences, representing 17.36% of all NPs and 12.6% of all difficulties. The other subcategories of NPs will not be investigated in the present thesis as they are less frequent, representing, altogether, 13.2% of NPs (35 occurrences) and 9.59% of all difficulties.

Single-word terms are the second most frequent category, with 34 occurrences (9.31%), followed by **numbers**, with 31 occurrences (8.49%), **proper names** (4.38%), culture-specific items, idioms and finally, phrasal verbs. These last three categories are the least frequent, totalling 19 occurrences all together, representing 5.20% of all difficulties. Some of the difficult items did not occur in some of the source speeches, such as culture-specific items in the first and second source speech and phrasal verbs in the second source speech. In the following analyses, I will focus on the most frequent difficult items, namely complex noun phrases, single-word terms, numbers and proper names.

In the difficulty profiles presented in Table 5, all the difficult items were taken into consideration, regardless of whether they were repeated (i.e. used several times in a speech) or not. This corresponds to an analysis of so-called tokens. However, for my thesis, I believe that it is more interesting to focus on **hapax legomena**[4]. They are defined as "linguistic expressions with only one attested occurrence and whose meaning is often, therefore, difficult to ascertain" (Bussman, 1996: 500). In other words, hapaxes are words that only occur once in a single text or corpus (Lardilleux & Lepage, 2007). If we analyzed tokens, the analysis could be biased as the interpreter trainee may use a different strategy when a given token is used again in the same speech. Therefore, recurring items were removed from the database. However, it is important to note that I relied on the first part of the definition given by Lardilleux & Lepage (2007), i.e. occurring only once "in a single text", because there are three items that occur in more than one speech. I decided to keep these three items because it can be interesting to see if trainees' outputs differ from one speech to another. Here are the three items: *vision* (single-word term in IN01 and IN02), *hearing* (single-word term in IN01, IN02 and IN04) and *Lord Reith* (proper name in IN02 and IN04). Table 6 shows the number of hapaxes for each difficulty in the four source speeches, with the recurring items removed.

[4] From the Greek *hapax legomenon* "which has been uttered once". The plural *hapaxes* will be used for convenience.

Table 6 – Difficulty profile of the source speeches (hapaxes)

	IN01	*IN02*	*IN03*	*IN04*	***Frequency***
Complex noun phrases	55	72	58	58	243 (73.19%)
A + N	28	20	27	20	95 (28.61%)
N + prep + N	11	27	15	17	70 (21.08%)
N + subclause	7	18	7	14	46 (13.85%)
N + N	2	3	5	6	16 (4.81%)
Adv + A + N	2	1	2	1	6 (1.81%)
other	2	3	1	0	6 (1.81%)
A + N + N	2	0	1	0	3 (0.90%)
N + N + N	1	0	0	0	1 (0.30%)
Single-word terms	9	5	3	10	27 (8.13%)
Numbers	4	10	6	7	27 (8.13%)
Proper names	1	4	7	3	15 (4.52%)
Culture-specific items	0	0	6	2	8 (2.41%)
Idioms	3	1	1	1	7 (2.11%)
Phrasal verbs	2	0	1	2	5 (1.51%)
Total	74	92	82	83	332 (100%)

Complex noun phrases are still, by far, the largest category of difficulties, with 243 occurrences, representing 73.19% of the difficulties. The three most frequent NP patterns, which will be the only patterns investigated here, are “A + N” (95 occurrences, 39.09% of NPs and 28.61% of all difficulties), “N + prep + N” (70

occurrences, 28.81% of NPs and 21.08% of all difficulties) and finally, "N + subclause" (46 occurrences, 18.93% of NPs and 13.85% of all difficulties). **Single-word terms** and **numbers** are the second most frequent difficulties, with 27 occurrences each (8.13%), followed by **proper names** (4.52%). My data analysis will focus on these four types of difficulties as they are the most frequent in LOCOSSI, and hence allow for generalizations.

3.4.2 Student interpretations: taxonomy of rendition types

Once I had identified all the IN difficulties, I went through all the students' interpretations to find out how they had handled them. The majority of interpreters' utterances can be considered as reformulations of prior original utterances and can therefore be named renditions (Wadensjö, 1998: 106). By definition, a rendition is "a stretch of text corresponding to an utterance voiced by an interpreter" and is "a second, and thus, in at least some respect, a new version of an original" (ibid.). Interpreters' renditions can relate to originals in different ways, which may form a basis for the classification of renditions into sub-categories. The present thesis is based on Wadensjö's classification of renditions (1998: 106-108), to which I added two categories, namely divergent rendition (Amato & Mack, 2011) and innovative rendition. Table 7 provides an overview of the taxonomy used, together with examples taken from LOCOSSI.

Table 7 – Taxonomy of renditions, based on Wadensjö (1998) and Amato & Mack (2011)

Type of rendition	Definition	Example from LOCOSSI
Close rendition	The content found explicitly expressed in the rendition must be equally found in the preceding original	*letters of the Martian alphabet* > *des lettres de l'alphabet martien* (Item #36 – IN01_STU01)

Reduced rendition	A reduced rendition includes less explicitly expressed information than the preceding original utterance (in other words, the rendition is simpler)	*number of possible permutations and combinations of human brain activity* > *nombre de permutations dans l'activité neuronale humaine* (Item #56 – IN02_STU03)
Expanded rendition	An expanded rendition includes more explicitly expressed information than the preceding original utterance	*several ways of studying the brain* > *plusieurs façons de s'y prendre dans l'étude du cerveau* (Item #86 – IN01_STU06)
Substituted rendition	It consists of a combination of an expanded and a reduced rendition	*huge site of temples* > *site où se trouvent des temples* (Item #31 – IN03_STU06)
Summarized rendition	It is a text that corresponds to two or more prior originals	*"parietal lobe of the brain"– "both sides of the head"* > *lobe pariétal de chaque coté du cerveau* (Item #79-#80 – IN01_STU07)
Divergent rendition	A divergent rendition contains elements that differ from what the speaker actually said, the output is not equivalent to the input	*four of the smartest people that I've ever met* > *parmi les personnes intelligentes que j'ai rencontrées aujourd'hui (vs. quatre des personnes les plus intelligentes que j'ai jamais rencontrées)* (Item #23 – IN04_STU04)
Innovative rendition	An innovative rendition contains elements (lexical	*epitome of feminine sensuality* >

	items, syntactic structures, etc.) that are not attested in the target language	*épitome de la féminité (vs. l'exemple parfait de la sensualité féminine)* (Item #85 – IN03_STU05)
Zero rendition	The input is left uninterpreted	*Rembrandt* > / (Item #82 IN03_STU04)

The different types of renditions will allow me to identify which items are the most difficult for interpreter trainees. On the one hand, a close rendition is considered as correct, while, on the other, a divergent rendition usually means that a mistake has been made. The other types of renditions are, however, much harder to judge, but some of them correspond to Gile's interpreting strategies and tactics (1995: 129-135). The strategies of *explicitation* and *paraphrasis* are equivalent to the *expanded* rendition. To the *reduced rendition* may correspond two other strategies, namely *simplification* (not rendering all the elements) and the *restitution on a higher level of abstraction*. Finally, the *zero rendition* may be equivalent to the *omission* (a conscious omission when the interpreter does not have enough time to process the information and does not recall what the speaker said).

4 Results and discussion

Before delving deeper into the interpreting difficulties investigated in the present case study, we will first take a closer look at some of the features of the source speeches and interpretations under scrutiny in Section 4.1, namely delivery rate, lexical density, and expansion rate. The idea here is to get a first general overview of all source and target speeches. These preliminary investigations will make it possible to outline some tendencies.

Then, I will analyse the database in order to answer my two research questions. Section 4.2 presents the quantitative analysis that will highlight the most challenging linguistic item for interpreter trainees among four different items, namely complex noun phrases (three sub-patterns), single-word terms, proper names, and numbers. Then, a longitudinal analysis of the data will be presented in Section 4.3 and will answer the second research question. In fact, in this section, proportions of close and divergent renditions are given for each interpreting task to see if there is some kind of improvement across time and across students.

Finally, Section 4.3 focuses on a case study concerning zero renditions. This type of rendition is really interesting to investigate as omissions can be very hard to judge (justified or unjustified). This case study focuses on the performances of two students.

4.1 Global features

The first global feature to be presented and analysed is delivery rate (Section 4.1.1). Then, Section 4.1.2 focuses on lexical density while Section 4.1.3 gives an analysis of the third global feature, expansion rate. Finally, Section 4.1.4 sums up the trends outlined in the previous sections.

4.1.1 Delivery rate

The delivery rate, also called speech rate, is measured as the number of words per minute (Plevoets & Defrancq, 2018). It is an important factor to establish intelligibility and clarity (Rodero, 2012). Here are some examples of average speech rates: a comfortable presentation has a rate between 100 and 150 words per minute (w/m) while radio hosts utter between 150 and 160 w/m and sports commentators between 250 and 400 w/m. As for interpreting, interpreters think that a delivery rate between 100 and 120 English w/m is optimal for speeches that are not read out from a written text (Li, 2010). However, this figure can change according to the speech type. Here is the comment from the message board of the International Association of Conference Interpreters (AIIC):

> We all know that a speaker speaking at about 100 to 120 words per minute is perfectly acceptable. However, there are exceptions to this: dense originals without much redundancy. Such speeches may seem excessively fast even if presented at 120 words per minute. Normally, however, any original exceeding 140 words per minute is fast. I have clocked speakers at 180 words per minute (Communicate 1999).

Generally speaking, the delivery rate is considered as slow if it is lower than 130 w/m, average between 130 and 160 w/m and fast over 160 w/m. Studies show that the delivery rate has a direct influence on interpreting quality, and accuracy decreases as the speech gets faster. In Gerver's study (1971), we see that the number of incorrectly translated words is caused by an increased delivery rate. Galli (1990) states that speech rate correlates with an increased number of omissions and mistakes, whereas Chernov (2004) says that the interpreter's delivery rate does not increase proportionally with the speaker's.

Table 8 – Overview of IN delivery rates

Source text	Number of words	Duration	Delivery rate (in words/minute)
IN01	1,378	0'28"-14'32" = 14'04"	97.96 w/m
IN02	1,343	0'05"-13'42" = 13'37"	98.63 w/m
IN03	1,487	0'00"-13'45" = 13'45"	108.14 w/m
IN04	1,414	0'36"-13'28" = 12'52"	109.9 w/m
IN average	1,405.5	13'34"	103.66 w/m

Table 8 displays the total number of words, duration and delivery rate of each source speech included in the corpus. All source speeches were read out by the same lecturer under the same conditions. As can be seen in Table 8, the delivery rate is slow for all IN (less than 130 words per minute). Interestingly, the delivery rate of each IN is higher than the preceding one. It is important to remember that the first IN was used on November 3rd, the second on November 10th, the third on November 17th and the last one on December 8th, 2016. I can thus conclude that the lecturer gradually increased her delivery rate (whether consciously or unconsciously), thereby increasing the difficulty of the interpreting tasks to be performed by the students.

Table 9 – Overview of STU delivery rates

	IN01 **97.96 w/m**	IN02 **98.63 w/m**	IN03 **108.14 w/m**	IN04 **109.9 w/m**	Average **103.66 w/m**
STU01	102.91 w/m	106.70 w/m	118.83 w/m	110.63 w/m	109.77 w/m
STU02	97.18 w/m	96.85 w/m	109.72 w/m	109.10 w/m	103.21 w/m
STU03	/	106.87 w/m	105.78 w/m	/	106.32 w/m
STU04	82.24 w/m	/	104.44 w/m	102.31 w/m	97.66 w/m
STU05	/	/	118.34 w/m	114 w/m	116.17 w/m
STU06	/	79.76 w/m	99.76 w/m	95.38 w/m	91.63 w/m
STU07	107.58 w/m	98.45 w/m	/	96.38 w/m	100.80 w/m
STU08	90.18 w/m	/	/	/	90.18 w/m
Average	96.02 w/m	97.73 w/m	109.48 w/m	104.63 w/m	**101.97 w/m**

Students' delivery rates are displayed in Table 9. As can be seen in Table 9, there are striking differences across interpreting tasks and across students. Generally speaking, Table 9 demonstrates that average STU delivery rate is lower than the speaker's (101.97 w/m vs. 103.66 w/m). However, if we take a closer look at the table, we see that this is not the case for each interpreting task. Out of the four interpreting tasks, one stands out. Indeed, the speaker's delivery rate for IN03 (108.14 w/m) is lower than the average STU delivery rate (109.48 w/m). We also see that, as Chernov says (2004), the students' delivery rate does not proportionally increase with the speaker's. IN03 delivery rate is lower than IN04 delivery rate (108.14 w/m < 109.90 w/m) but STU average delivery rate is higher for IN03 than IN04 (109.48 w/m > 104.63 w/m).

There are also marked differences across students. First, the output delivery rates are considered as slow because they are all lower than 130 w/m. Four students,

namely STU01 (109.77 w/m), STU02 (103.21 w/m), STU03 (106.32 w/m) and STU05 (116.17 w/m), have an average delivery rate which is markedly higher than the average IN delivery rate (101.97 w/m). Two other students, STU04 and STU08, systematically have a lower average delivery rate. It is also interesting to see that STU01's and STU05's delivery rates are higher than the average IN delivery rate for each interpreting task.

4.1.2 Lexical density

Halliday (1985: 62) says that lexical density is "the number of lexical items, as a proportion of the number of running words" and adds that "lexical density if the kind of complexity that is typical of written language". Johansson (2008: 61) defines lexical density as "a measure of the proportion of lexical items (i.e. nouns, verbs, adjectives and some adverbs) in the text". Lower lexical density suggests that the text is less informative. Information is diluted with more function words. Between two texts of the same length, the one containing more lexical words would be considered more informative.

Before calculating the lexical density of all speeches, it is important to define lexical and grammatical words. There is first a distinction to be made between **closed-class** and **open-class** parts of speech (Jurafsky & Martin, 2004: 3): "Closed classes are those that have relatively fixed membership. For example, prepositions are a closed class because there is a fixed set of them in English; new prepositions are rarely coined. By contrast nouns and verbs are open classes because new nouns and verbs are continually coined or borrowed from other languages". Closed-class words are **grammatical words**, whereas open-class words are **lexical words**. Grammatical words mainly comprise prepositions, determiners, pronouns, auxiliaries, conjunctions, particles, numerals, interjections, negatives, greetings, and politeness markers. The main groups of lexical words are nouns, adjectives, verbs (whether lexical or grammatical) and adverbs.

Laviosa (1998) worked on lexical density in TEC, the *Translational English Corpus*. This corpus is made of translated narrative prose (from a number of languages into English) and of original English narrative texts. She found that translated texts have a significantly lower number of content words versus grammatical words, and their lexical density is thus lower (1998: 563). Sandrelli & Bendazzoli investigated EPIC to see if this was also accurate for interpreted speeches. The effect noted by Laviosa (1998) was not confirmed in their study as the variation between interpreted and original speeches is very little. Kajzer-Wietrzny (2005) also showed that there was no consistent difference in the informativeness of interpreted and original speeches, as lexical densities were very similar (57% for original English and 56% for interpreted French).

As it is shown in Appendix 3, I have distinguished between lexical, grammatical and irrelevant POS-tags. Three POS-tags were not considered in the lexical density scores reported in Table 10: LS (list markers), SENT (sentence-break punctuation) and SYM (symbol) because they were irrelevant to the calculation of lexical density.

I will first take a look at the lexical density of the four source speeches. Table 10 shows that all source speeches are rather dense (between 52% and 58%). The lexical density scores are higher than those obtained in spoken corpora. Ure (1971) argues that most spoken texts have a lexical density of under 40%, while a large majority of written texts have a lexical density of 40% or higher. This seems to suggest that the source speeches included in LOCOSSI are closer to writing than speech.

Table 10 – IN lexical density

IN	Number of words	Number of lexical words	Lexical density
IN01	1,378	772	56.02%
IN02	1,343	718	53.46%
IN03	1,487	860	57.83%
IN04	1,414	740	52.33%
Average	1,405.5	772.5	54.96%

In order to check this, I have computed the lexical density in two other spoken corpora: English speeches in EPTIC and the spoken demographic part of the BNC. The *European Parliament Translation and Interpreting Corpus* (EPTIC) is an intermodal, comparable and parallel corpus, which comprises speeches delivered at the EP with their official interpretations and translations (Bernardini et al., 2016). I relied on the English source speeches of EPTIC (23,549 words, see Table 11). The second corpus used for the comparison is the spoken_demographic subcorpus of the *British National Corpus* (BNC). The BNC, which is a collaboration between commercial and academic partners, is a 100-million-word collection of samples of written and spoken language from a wide range of sources. It aims to represent a wide cross-section of British English from the later part of the 20th century (Leech, 1992). The written part represents 90% of the BNC while the spoken part represents 10% of it. The latter consists of transcriptions of informal conversations and spoken language gathered in different contexts. The spoken_demographic subcorpus contains 4,190,072 words.

Table 11 – Lexical density across spoken corpora

Subcorpus	Number of words	Number of lexical words	Lexical density
EPTIC-st-in-en	23,549	16,088	68.32%
BNC_spoken_demographic	4,657,760	2,256,866	48.45%
IN	5,622	3,090	54.96%

As shown in Table 11, lexical density is much higher in EPTIC than in the BNC_spoken_demographic subcorpus, which is quite logical given the different genres represented in each corpus. In the first case, the corpus is made of speeches delivered by Members of the European Parliament while the second consists of everyday conversations that were collected in contexts such as radio shows or phone-ins. Interestingly, we see that the source speeches included in the corpus display a lexical density higher than everyday conversations but lower than parliamentary debates.

Two POS-tags were not taken into consideration in the lexical density scores reported for the French outputs: SENT (sentence-break punctuation) and SYM (symbol) because they were irrelevant to the calculation of lexical density. Table 12 shows the lexical density of all interpreting tasks and all students' outputs.

Table 12 – Lexical density (LD) across inputs and outputs in LOCOSSI

	IN01	IN02	IN03	IN04	**Average LD score**
	56.02%	53.46%	57.83%	52.33%	54.96%
STU01	54.72%	52.19%	52.27%	52.55%	52.69%
STU02	52.89%	51.81%	50.56%	51.36%	51.30%
STU03	/	52.71%	53.39%	/	52.28%
STU04	54.79%	/	52.72%	52.55%	52.82%
STU05	/	/	53.38%	51.88%	52.15%
STU06	/	52.21%	50.76%	53.03%	51.47%
STU07	53.54%	53.45%	/	53.21%	52.76%
STU08	53.35%	/	/	/	53.35%
Average LD score	53.84%	52.48%	52.19%	52.39%	**52.35%**

First, there is no significant difference between the STU average (52.35%) and the IN average (54.96%) lexical density score, which means that students' interpretations are as informative as the original speeches. This confirms Kajzer-Wietrzny's (2005) and Sandrelli & Bendazzoli (2005) findings.

Table 13 – Number of running words in all IN and STU speeches

	IN01 1,378	IN02 1,343	IN03 1,487	IN04 1,414
STU01	1,451	1,460	1,632	1,429
STU02	1,367	1,322	1,505	1,400
STU03	/	1,457	1,458	/
STU04	1,170	/	1,436	1,313
STU05	/	/	1,641	1,463
STU06	/	1,086	1,375	1,205
STU07	1,524	1,334	/	1,216
STU08	1,267	/	/	/
Average number of running words	1,355.8	1,331.8	1,507.83	1,337.67

However, it appears that, in most cases, interpretations are less dense than the corresponding source speech (an exception is IN04, where no difference can be observed). Even though there are striking differences in terms of the total number of words produced by students for the same IN (see Table 13), a close inspection of Table 12 reveals that the lexical density scores are surprisingly similar across students. For example, STU04 produced 1,170 words to interpret IN01 while STU07 produced 1,524 words to interpret the same text (23.22% more words). The higher lexical density of STU04's output shows that STU04 said fewer words but that a large share of these words were content words.

4.1.3 Expansion rate

In interpreting, many studies investigated “compression” (Chernov, 1987, Sdobnikov, 2016) and proposed typologies of compression. It consists of any shortening of the linguistic form used to express a notion or an idea in simultaneous interpreting (Sdobnikov, 2017). Compression can occur at different levels of language, such as syllables, words, syntax, context, or meaning. According to Chernov, compression is a possible mechanism and can be used to avoid “the linguistic redundancy in the thematic component of the discourse” (2004: 113).

However, there is another mechanism, widely under-researched so far, called decompression. The process of decompression, or “expansion”, is the “result of a longer/expanded idea vs. the probable compressed version we might use in the normal conditions of interpretation” (Sdobnikov, 2017: 411). In other words, the mechanism is opposite to compression. It is important to note that this phenomenon is often considered as an interpreting strategy as Donato (2003: 107) places it in the reformulation strategies, Liontou (2011: 41) among target-text conditioned strategies and Ricardi (2005: 765) puts it in the production strategies.

In this case study, I decided to use this notion of expansion and investigate the expansion rate of student interpretations in order to determine whether the interpretations were longer than the corresponding original speeches. If the expansion rate is higher than 1, it means the interpretation is longer (in terms of the total number of words used) than the original. If it is lower than 1, then the interpretation is shorter than the original. The expansion rate is calculated by dividing the number of words of an interpretation by the number of words of its source speech.

Table 14 – Expansion rates in LOCOSSI

	IN01	IN02	IN03	IN04	Average
STU01	1,451/1,378 = 1.053	1,460/1,343 = 1.087	1,632/1,487 = 1.097	1,429/1,414 = 1.011	1.062
STU02	1,367/1,378 = 0.992	1,322/1,343 = 0.984	1,505/1,487 = 1.012	1,400/1,414 = 0.990	0.995
STU03	/	1,457/1,343 = 1.085	1,458/1,487 = 0.980	/	1.033
STU04	1,170/1,378 = 0.849	/	1,436/1,487 = 0.966	1,313/1,414 = 0.928	0.914
STU05	/	/	1,641/1,487 = 1.104	1,463/1,414 = 1.035	1.069
STU06	/	1,086/1,343 = 0.801	1,375/1,487 = 0.925	1,205/1,414 = 0.854	0.862
STU07	1,524/1,378 = 1.106	1,334/1,343 = 0.993	/	1,216/1,414 = 0.860	0.986
STU08	1,267/1,378 = 0.919	/	/	/	0.919
Average	0.984	0.990	1.014	0.946	**0.980**

As shown in Table 14, the overall average expansion rate in LOCOSSI is lower than 1 (value = 0.980), which means that interpreter trainees used the mechanism of compression. This corpus presents little evidence of compression, which confirms Chernov's assumption (2004). However, Chernov's theoretical framework should not be taken for granted as other studies did not show any evidence of compression (Iacovoni, 2010).

In general, interpreter trainees used fewer words in their interpretations. This is especially striking in STU04 and STU06's interpretations. If we look at the table somewhat closer, we see that there are many differences across students and not so many across interpreting tasks. IN03 stands out as its STU average expansion

rate is the only one being higher than 1 but, in general, expansion rate scores are very similar across interpreting tasks.

Linking the delivery rate scores and the expansion rate scores can allow to outline some tendencies across students (as we said, lexical density scores are surprisingly similar across students). It is interesting to see that the majority of students who have an average delivery rate markedly higher than the IN average delivery rate, namely STU01, STU03, STU05, also have an average expansion rate higher than 1. In the same vein, the lowest delivery rate for IN01 and IN02 was respectively STU04's score (82.24 w/m) and STU06's score (79.76 w/m) and the lowest expansion rate for IN01 and IN02 was respectively STU04's rate (0.849) and STU06's rate (0.801).

4.1.4 Interim conclusion

Summing up the trends outlined in the previous sections, we can say that:

DELIVERY RATE (DR):

- Slow average delivery rate (< 130 w/m): 101.97 w/m for STU and 103.66 w/m for IN
- Average STU delivery rate < average IN delivery rate
- STU delivery rate does not increase proportionally with the speaker's delivery rate (Chernov, 2004)
- *Student profiles:* STU01, STU03 and STU05 have the highest average delivery rate, whereas STU06 and STU08 have the lowest

LEXICAL DENSITY (LD):

- STU interpretations are as informative as source speeches (average LD score respectively being 52.35% and 54.96%) as the difference is not significant (Kajzer-Wietrzny, 2005; Sandrelli & Bendazzoli, 2005)
- Lexical density scores are surprisingly almost similar across students, but, despite having the lowest DR and one of the lowest ER, STU08 has the highest LD score (53.35%), which means that STU08 used fewer words but the interpretations were denser than the others

EXPANSION RATE (ER):

- Average expansion rate is lower than 1, little evidence of compression (Chernov, 2004)
- Same tendencies across students than the ones observed for delivery rate: if the STU delivery rate is higher than the speaker's, then the STU expansion rate is higher than 1
- *Student profiles*: STU01, STU03 and STU05 are the only students to have a positive average expansion rate (> 1), whereas STU06 and STU08 have a negative average expansion rate (< 1)

4.2 Data analysis

In the remainder of this chapter, which is mainly devoted to the way students handled potential difficulties (four different linguistic items), I will answer my two research questions and try to determine whether the trends outlined in Section 4.1.4 can be linked to students' performances.

4.2.1 Introduction

In this section, I address the following research question:

> Which source linguistic items are the most challenging for interpreter trainees working from English into French? In tackling this question, I also aim to show how it is possible to determine that a specific item is challenging or difficult for an interpreter.

Operationalizing interpreting quality as a low number of errors (i.e. divergent renditions), I would like to address a second research question:

> Does the quality of the trainees' interpretations improve across time? How can we measure that improvement? Do all trainees improve across time?

4.2.2 Quantitative analysis

In order to answer my first research question, it is important to analyse the percentage of close and divergent renditions. A close rendition means that the content of the trainees' output closely corresponds to what is found in the original. A divergent rendition, by contrast, is a case of interpreting error because the output is not equal to the input (Wadensjö, 1998: 106-108). However, it would also be interesting to examine the other types of rendition as they may shed light on interpreting strategies typical of given items (e.g. a high percentage of zero

renditions). The present analysis only focuses on the four most frequent linguistic items in the source speeches, i.e. complex noun phrases, with 243 occurrences, single-word terms, numbers and proper nouns, occurring 27, 27 and 15 times respectively. The number of occurrences in the TT is given in Table 15. As can be seen, the study is based on the analysis of 1,536 TT occurrences in total.

Table 15 – Occurrences of the 4 most frequent linguistic items in LOCOSSI

Linguistic item	**IN**	**STU**
Complex noun phrases	243	1,155
A + N	95	522
N + prep + N	70	382
N + subclause	46	251
Single-word terms	27	148
Numbers	27	148
Proper nouns	15	85
Total	312	1,536

4.2.2.1 Complex noun phrases

Table 16 – Percentage of each type of rendition for NPs

Type of rendition	**NPs**	**%**
Close rendition	587	50.84%
Divergent rendition	214	18.56%
Reduced rendition	194	16.80%
Zero rendition	105	9.07%
Expanded rendition	31	2.63%
Substituted rendition	11	0.99%
Summarized rendition	8	0.70%
Innovative rendition	5	0.40%
Total	**1155**	**100%**

Table 16 shows the percentage of each type of rendition for NPs. Half of complex noun phrases are correctly interpreted (50.84%), see *examples 42 and 43*, while a non-negligible number of NPs (almost one NP out of five, 18.56%) is wrongly interpreted (*examples 44 and 45*).

(42) *across-the-board reduction of their cognitive capacities* > réduction totale de leurs capacités cognitives (Item #91 - IN02_STU01)

(43) *average female form* > *forme d'une femme normale* (Item #86 - IN03_STU03)

(44) *mere speck of dust revolving around the sun* > *une simple planète parmi un système entier* (Item #9 – IN02_STU01)

(45) *sickle cell anaemia* > *envoi d'e-mails séquentiels* (Item #3 – IN01_STU07)

As highlighted by Bartlomiejczyk (2006) and Kader & Seubert (2015), a useful strategy to interpret complex noun phrases is compression (also called condensing), which corresponds to reduced renditions in Wadensjö taxonomy. This phenomenon (whether conscious or not) is quite frequent in the dataset, with 16.8% of all renditions of NPs.

(46) *brain ability* > *capacité* (Item #59 – IN01_STU04)

(47) *soft-boiled egg* > *oeuf* (Item #9 – IN02_STU02)

Zero renditions (omissions) are quite infrequent. They are the fourth most frequent type of rendition (9.07%). Omissions are not necessarily wrong. As shown by Gile (1995), an omission can be justified or unjustified. A justified, or conscious, omission can be made when the interpreter does not have enough time to process the information and does not recall what the speaker said. In some rare cases, students decided to explicitate the noun phrase (2.63%).

(48) *wouldn't give me a penny for it* > *vous ne me donneriez absolument rien pour cela, ça ne vaut pas un centime* (Item #47 – IN03_STU05)

(49) *pleasing effect in the brain* > *effet qui soit positif pour le cerveau* (Item #50 – IN03_STU01)

Finally, it is interesting to note that, in some cases, students are heavily influenced by the source language (here English) and use words that are not attested in the target language (so-called innovative renditions). However, this type of rendition is extremely rare in the present dataset (0.40%).

(50) *epitome of feminine sensuality* > ***épitome*** *de la sensualité* (Item #85 – IN03_STU05)

(51) (10) *cross-modal synesthetic abstraction* > *abstraction* ***synesthésiste*** (Item #49 – IN01_STU02)

First pattern: A + N

The first pattern to be analysed is "A + N", with 95 occurrences in the source speeches and 522 in the students' interpretations. It includes three sub-patterns:

- Adjective + noun: ***white*** *noise,* ***perfect*** *prosthetic*
- Adjective + adjective + noun: ***round curvy*** *shape,* ***long white*** *thing*
- Adverb + adjective + noun: ***very small*** *lesion,* ***quite common*** *knowledge*

Table 17 – Renditions for the "A + N" pattern

	STU01	STU02	STU03	STU04	STU05	STU06	STU07	STU08	**Total**
Close	53/95	52/95	26/47	40/75	32/47	34/67	34/68	14/28	54.75%
Div.	19/95	11/95	4/47	13/75	7/47	11/67	10/68	4/28	14.72%
Red.	14/95	16/95	6/47	7/75	5/47	9/67	10/68	5/28	13.79%
Zero	5/95	14/95	7/47	14/75	2/47	10/67	10/68	4/28	12.72%
Exp.	2/95	/	3/47	1/75	/	1/67	3/68	1/28	2.41%
Summ.	2/95	/	1/47	/	1/47	2/67	1/68	/	1.35%
Subst.	/	1/95	/	/	/	/	/	/	0.13%
Innov.	/	1/95	/	/	/	/	/	/	0.13%

As Table 17 shows, close renditions account for slightly more than half of the occurrences, with 54.75% of the occurrences (see examples 52 and 53), and divergent renditions are the second most frequent type of rendition, with 14.72% of

the occurrences (see examples 54 and 55). This percentage is lower than the average percentage of divergent renditions for all complex noun phrases considered (14.72% compared to 18.56%).

(52) *common denominator* > *dénominateur commun* (Item #73 – IN01_STU08)
(53) *basic anatomy* > *anatomie de base* (Item #64 – IN02_STU01)
(54) *artificial knees* > *genoux extérieurs* (Item #18 – IN04_STU07)
(55) *hidden agenda* > *agenda caché* (Item #4 – IN01_STU07)

The two next most frequent types of rendition closely follow divergent renditions and are reduced renditions (13.79%) and zero renditions (12.72%). The latter percentage is higher than the average percentage of zero renditions for all NPs considered (12.72% to 9.07%). The least frequent types of renditions are expanded (2.41%), summarized (1.35%), substituted (0.13%), and innovative renditions (0.13%).

It is worth noting individual variations as well. Concerning zero renditions, there is, on the one hand, a significant difference in their frequency between STU01 and STU02 (who interpreted the same number of “A + N” NPs). The percentage of zero renditions for STU02 is 14.74% (5 omissions) while STU01’s percentage of zero renditions is 5.26% (14 omissions). On the other hand, STU05 has the lowest percentage of zero renditions (4.30%), while STU04 has the highest (18.67%).

There are other striking facts across students. Every student’s percentage of close rendition is at least 50% of all occurrences. However, STU05 stands out as close renditions account for 68.09% of all occurrences, which means that more than two-thirds of “A + N” NPs were correctly translated. In addition, STU02 is the only student to have invented a word. As the example below shows, STU02 did not know the translation of “synesthetic” and should have translated it as “synesthésique”.

(56) cross-modal synesthetic abstraction > abstraction **synesthésiste** (Iteam #49 – IN01_STU02)

Second pattern: N + prep + N

The second pattern to be analysed is "N + prep + N", with 70 occurrences in the source speeches and 382 in the students' interpretations. Five different prepositions are used to form that kind of NP: *of*, *for*, *into*, *about*, and *with*. The preposition *of* is by far the most frequently used preposition, representing 88.57% of all prepositions. Table 18 shows the different types of rendition for this pattern.

(57) shape **of** the rectangle (Item #67 – IN03)
(58) capacity **for** abstraction (Item #77 – IN01)
(59) insights **into** the human brain (Item #25 – IN02)
(60) interesting thing **about** prosthetics (Item #16 – IN04)
(61) someone **with** a good hearing (Item #44 – IN04)

Table 18 – Renditions for the "N + prep + N" pattern

	STU01	STU02	STU03	STU04	STU05	STU06	STU07	STU08	**Total**
Close	33/70	45/70	23/42	18/43	22/32	19/59	31/55	6/11	52.49%
Div.	17/70	8/70	11/42	10/43	3/32	18/59	4/55	/	16.54%
Red.	12/70	9/70	2/42	6/43	1/32	12/59	13/55	3/11	15.39%
Zero	3/70	6/70	2/42	7/43	3/32	7/59	5/55	2/11	10.30%
Exp.	5/70	2/70	2/42	1/43	1/32	1/59	/	/	2.74%
Innov.	/	/	1/42	/	2/32	/	/	/	1.08%
Summ.	/	/	1/42	/	/	/	2/55	/	0.75%
Subst.	/	/	/	1/43	/	2/59	/	/	0.71%

When we compare the percentages for the different types of rendition for this pattern and the first pattern analysed, there are not significant differences. The five most frequent types of rendition are the same: close (52.49%), divergent (16.54%), reduced (15.39%), zero (10.30%) and expanded (2.74%) renditions. Then come innovative, summarized and substituted renditions but they are quite rare as their respective percentages are 1.08%, 0.75% and 0.71%. I will focus on summarized

renditions, which show the use of compression, an interpreting strategy to limit the redundancy of elements in a discourse (Chernov, 2004).

(62) *three-dimensional representation of the spatial layout of the external world - three-dimensional representation of your own body > créer une representation 3D du monde extérieur mais également de notre* corps (Items #81 & #82 – IN02_STU03)

(63) *parietal lobe of the brain - both sides of the head > lobe partiétal de chaque côté du cerveau* (Items #79 & #80 – IN02_STU07)

If we take a closer look at the percentage of close renditions of each STU, we note that three students did not interpret half of this type of NP correctly. Indeed, STU01 only interpreted 47.14% of them correctly, while STU04 and STU06 respectively interpreted 41.86% and 32.20% correctly. It is clear that STU06 stands out as the percentages of close and divergent renditions are almost equivalent (32.20% vs. 30.51%). This pattern seems to be quite difficult for that student as it is the highest percentage of divergent renditions. On the contrary, it seems that STU05 deals with this pattern of NPs well as close renditions account for 68.75% and divergent renditions only for 9.37% of all occurrences (half of the average percentage of divergent rendition for all NPs). However, it is quite surprising to see that STU05 also invented two words in French: *la pose* and *l'épitome*.

(64) *very epitome of feminine sensuality, grace, poise, dignity* > ***l'épitome*** *de la sexualité féminine, la grâce, la* ***pose****, la dignité* (Item #36 – IN03_STU05)

(65) ***epitome*** *of feminine sensuality* > ***épitome*** *de la féminité* (Item #85 – IN02_STU03)

Finally, I would like to focus on one specific item that was wrongly interpreted by all students: *epitome of feminine sensuality* (Item #85 – IN03). This item was interpreted as *idées de sensualité féminine*, *féminité*, or *essence de la féminité,* among other interpretations.

Third pattern: N + subclause

The last pattern to be analysed is "N + subclause", with 46 occurrences in the source speeches and 251 in the students' interpretations. There are eight different types of items that help forming this pattern: *that*, *which*, *who*, *where*, *why*, *gerund*, *past participle*. The complementizer *that* is the most frequent item used to form this pattern (50% of all occurrences).

(66) *iron hand **that** belonged to one of the German counts* (Item #6 – IN04)

(67) *lecture **which** is the most speculative of the series* (Item #1 – IN03)

(68) *Victorian English men **who** first saw these pictures* (Item #37 – IN03)

(69) *places **where** exchange of information occurs* (Item #55 – IN02)

(70) *reason **why** these are so interesting* (Item #34 – IN04)

(71) photons **hitting** the retina (Item #71 – IN01)

(72) *cultural boundaries **applied** everywhere in the* world (Item #57 – IN03)

Table 19 – Renditions for the "N + subclause" pattern

	STU01	STU02	STU03	STU04	STU05	STU06	STU07	STU08	**Total**
Close	20/46	23/46	12/25	9/28	9/21	12/39	17/39	5/7	45.28%
Div.	11/46	12/46	4/25	11/28	6/21	14/39	10/39	/	24.42%
Red.	13/46	7/46	7/25	2/28	4/21	8/39	9/39	2/7	21.23%
Zero	/	/	/	3/28	1/21	5/39	2/39	/	4.18%
Exp.	2/46	3/46	1/25	2/28	/	/	/	/	2.75%
Subst.	/	1/46	1/25	1/28	1/21	/	1/39	/	2.13%

As it can be seen in Table 19, this pattern is markedly different from the other patterns, with one fourth of divergent renditions (24.42%) and a fifth of reduced renditions (21.23%). These two percentages are much higher than the average percentages, being respectively 18.56% and 16.80%. As this pattern is quite frequent in the source speeches (46 occurrences), syntactic complexity is high. It is measured as the number of subordinate clauses per utterance length (Plevoets

& Defrancq, 2018). Close renditions also drop and reach 45.28% of all occurrences, which is the lowest percentage of close renditions among the three patterns. Given the high syntactic complexity, which is a well-known predictor of high cognitive load (Gile, 2008; Chmiel & Mazur, 2013), we can thus conclude that this is clearly the most challenging pattern for students.

In addition, it is important to note that, for two students, namely STU04 and STU06, the proportion of divergent renditions (39.29% and 35.90%) is more important than the proportion of close renditions (32.14% and 30.77%). Finally, zero renditions surprisingly have very low percentages (4.18% compared to 9.07%, the average percentage for all NPs). It appears that students did not omit the items but rather condensed them (high percentage of reduced renditions).

4.2.2.2 Single-word terms

There are 27 occurrences of single word-terms in the source speeches and 148 in the students' interpretations. As can be seen in Table 20, there is a really high average percentage of close renditions, showing that almost three-quarters of terms were correctly translated.

Table 20 – Renditions for single-word terms

	STU01	STU02	STU03	STU04	STU05	STU06	STU07	STU08	**Total**
Close	21/27	17/27	6/8	14/22	9/13	16/18	17/24	6/9	71.87%
Zero	1/27	2/27	1/8	3/22	3/13	/	3/24	1/9	10.49%
Div.	4/27	6/27	/	4/22	/	2/18	1/24	1/9	10.20%
Red.	1/27	2/27	/	1/22	1/13	/	2/24	1/9	5.35%
Innov.	/	/	/	/	/	/	1/24	/	0.52%

Zero renditions are the second most frequent type of rendition, with 10.49%, which is quite logical given that single-word terms are short and refer to a specific concept or field (Khan, 2016). If the student does not have that knowledge, then an omission can be made.

(73) *amoeba* (Item #32 – IN01): omitted by four students

(74) *titanium* (Item #58 – IN04): omitted by three students

It is really surprising to see that 5.35% of single-terms were condensed. In fact, some items were generalized or simplified as they are made of more than one element. Finally, there is one innovative rendition (see example 77), probably due to a slip of the tongue. Nevertheless, the word does not exist in the target language (French).

(75) *cross-connections* (Item #60 – IN01) > *ce genre de connexions* (STU01), *ces connexions* (STU04)

(76) *neuroscientists* (Item #6 – IN03) > *scientifiques* (STU02, STU05)

(77) *synapses* > *synapsis* (Item #54 - IN02_STU07)

Across students, STU02 stands out by having the highest percentage of divergent renditions (22.22%) and the lowest percentage of close renditions (62.96%). Single-word terms are then the most challenging for STU02. In addition, it is surprising to see the proportion of omitted single-word terms by STU05, reaching 23.07% of all items (these items being *virus*, *superhearing*, and *hearing*).

4.2.2.3 Numbers

The following linguistic item to be analysed is numbers, with 27 occurrences in the source speeches and 148 in the students' interpretations. Many studies say that the error rates can reach 40% to 70% for trainees (Korpal, 2016; Korpal & Stachowiak, 2018; Timarová, 2012) so I would expect such an error rate in this present study. Interpreters also use strategies such as reformulation (Jones, 2002;

Pinochi, 2009) or approximation (Kalina, 1992: 254) so I would expect the share of expanded and reduced renditions to be quite significant.

Table 21 – Renditions for numbers

	STU01	STU02	STU03	STU04	STU05	STU06	STU07	STU08	**Total**
Close	20/27	24/27	13/16	13/17	9/13	16/23	15/21	4/4	78.86%
Div.	3/27	3/27	1/16	1/17	2/13	3/23	3/21	/	9.63%
Zero	2/27	/	1/16	3/17	2/13	2/23	3/21	/	8.71%
Exp.	1/27	/	1/16	/	/	1/23	/	/	1.72%
Red.	1/27	/	/	/	/	1/23	/	/	1.01%

Table 21 does not confirm my hypotheses. Indeed, divergent renditions only account for 9.63% of all renditions, and expanded renditions represent 1.72% of them. However, it is important to shed light on the numbers analysed in this study. Most of them are expressed using one or two words and do not seem to be a problem for interpreter trainees.

(78) ***twelfth*** *century* (Item #34 – IN03)

(79) ***ten*** *laws of art* (Item #55 – IN03)

(80) ***four*** *of the smartest people* (Item #23 – IN04)

Students dealt with numbers very well as more than three-quarters of the occurrences were correctly translated (78.86%). I expected students to use two strategies, namely reformulation and approximation, but it was not the case at all since expanded and reduced renditions respectively account for 1.72% and 1.01% of all renditions. It is also interesting to see that STU03 and STU04 have the lowest percentage of divergent renditions among all students, with 6.25% and 5.88% of all renditions. Furthermore, STU02 (88.88%) and STU03 (81.25%) have the highest percentage of close renditions and, apart from STU08 (excluded of the

analysis as STU08 only interpreted four occurrences), they are the only two students to have a higher percentage of close renditions than the average for this linguistic item (78.86%).

4.2.2.4 Proper names

The last linguistic item under investigation is proper names, with 15 occurrences in the source speeches and 85 in the trainees' interpretations. These 15 proper names were rather challenging for interpreter trainees as the percentage of zero renditions is quite high (14.21%), as can be seen in Table 22. Hanaoka (2002) insists on the difficulty to retain the proper name in the short-term memory and then to recode it in the target language. This process often causes a cognitive load and leads to omissions, as it is the case here. Proper names such as *Rembrandt* (Item #82 – IN03) or *Snow* (Item #27 – IN02) were omitted by more than half the students. However, more than two-thirds of proper names were correctly translated (68.66%). Finally, Quini (1993) says that interpreters may risk mispronouncing the name if they do not know how the word is spelled beforehand or if they are unfamiliar to that proper name. This is what probably happened with the proper name *Hugh Herr* (Item #12 – IN04):

(81) *Hugh Herr* > *Frank Frank* (STU04 – divergent rendition)

(82) *Hugh Herr* > *June Herr* (STU06 – divergent rendition)

There are some striking facts across students. If we take a closer look at close renditions, we can see that STU06 has the lowest percentage of correctly translated proper names (42.86%). In addition, STU06 also has the highest percentage of wrongly translated items (28.57%), which means that proper names are a major difficulty for that student.

Table 22 – Renditions for proper names

	STU01	STU02	STU03	STU04	STU05	STU06	STU07	STU08	**Total**
Close	11/15	11/15	8/11	6/11	7/10	6/14	5/8	1/1	68.66%
Zero	3/15	/	2/11	1/11	2/10	3/14	2/8	/	14.21%
Div.	/	3/15	/	3/11	1/10	4/14	1/8	/	12.29%
Red.	1/15	1/15	1/11	1/11	/	1/14	/	/	4.82%

4.2.2.5 Interim conclusion

Table 23 – Overview of renditions for all studied linguistic items

	Close	Divergent	Zero	Reduced	Expanded	Substituted	Innovative	Summarized
Numbers	78.86%	9.63%	8.71%	1.01%	1.72%	/	/	/
Single-word terms	71.87%	10.20%	10.49%	5.35%	/	/	0.52%	/
Proper names	68.66%	12.29%	14.21%	4.82%	/	/	/	/
Complex noun phrases	50.84%	18.56%	9.07%	16.80%	2.63%	0.99%	0.40%	0.7%
A + N	54.75%	14.72%	12.72%	13.79%	2.41%	0.13%	0.13%	1.35%
N + prep + N	52.49%	16.54%	10.30%	15.39%	2.74%	0.71%	1.08%	0.75%
N + subclause	45.28%	24.42%	4.18%	21.23%	2.75%	2.13%	/	/
Average	67.56%	12.67%	10.62%	6.99%	1.09%	0.25%	0.23%	0.17%

Summing up the trends outlined in the previous section, we can say that:

Complex noun phrases:

- On a more general note, just over half of NPs are correctly translated (50.84%) and almost a fifth of them are wrongly translated (18.56%)
- Significant share of reduced renditions (16.80%) → use of interpreting strategies: compression
- Three sub-patterns:
    - **A + N**: over half of the occurrences correctly translated (54.75%)
        - **Zero renditions:** significant difference in the number of close renditions between STU01 and STU02 (5.26% vs. 14.74%)
    - **N + prep + N**: over half of the occurrences correctly translated (52.49%) and almost as many reduced and divergent renditions (15.39% vs. 16.54%)
        - **STU06** stands out as the percentage of close and divergent renditions is almost similar (32.20% vs. 30.51%)
    - **N + subclause**: very low percentage of close renditions (45.28%) and high percentage of divergent renditions (24.42%) and of reduced renditions (21.23%) → definitely the most challenging pattern for interpreter trainees as less than half of the occurrences were correctly translated
        - Two students stand out as their percentage of divergent renditions is higher than their percentage of close renditions: **STU04** and **STU06**

Single-word terms:

- High percentage of close renditions (71.87%) and low percentage of divergent renditions (10.20%) → manageable linguistic items and not

the most challenging linguistic items for interpreter trainees

- Two students stand out:
    - **STU02**: lowest percentage of close renditions (62.96%) and highest percentage of divergent renditions (22.22%) → least accurate performance
    - **STU06**: highest percentage of close renditions (88.89%) and lowest percentage of divergent renditions (11.11%) → most accurate performance

Numbers:

- Very high percentage of close renditions (78.86%) and low percentage of divergent renditions (9.63%) → most manageable linguistic item for interpreter trainees
- Two students stand out:
    - **STU04**: lowest percentage of close renditions (69.23%) and highest percentage of divergent renditions (15.38%) → least accurate performance
    - **STU02**: highest percentage of close renditions (88.89%) and quite low percentage of divergent renditions (11.11%, which is above the average percentage) → most accurate performance

Proper names:

- Quite high percentage of close renditions (68.66%), surprisingly high percentage of zero renditions (14.21%) → manageable linguistic item but potential source of high cognitive load due to the difficulty of retaining the proper name in the short-term memory
- Two students stand out:

- **STU06**: lowest percentage of close renditions (42.86%) and highest percentage of divergent renditions (28.57%) → least accurate performance
- **STU01**: highest percentage of close renditions (73.33%) and lowest percentage of divergent renditions (0%) → most accurate performance

As I said at the beginning of this section, it is important to analyse the percentage of close and divergent renditions in order to answer the first research question as these two types of rendition give clear evidence on the quality of the interpretation.

The different analyses carried out in the present study demonstrate that **complex noun phrases** are the most challenging linguistic items for interpreter trainees working from English into French. As can be seen in Table 23, **complex noun phrases** have the lowest percentage of close renditions and the highest percentage of divergent renditions among the four investigated linguistic items. It is possible to go even further and underline the fact that there is a particular NP sub-pattern, which seems very difficult → "N + subclause", i.e. any noun phrase followed by a subclause. Close renditions for this particular sub-pattern account for less than 50% of all renditions, which means that less than half of noun phrases followed by a subclause were correctly translated.

4.2.3 Longitudinal analysis

In order to answer the second research question, I will focus on the students who carried out at least three interpreting tasks, namely STU01, STU02, STU04, STU06 and STU07. I will only take into consideration close and divergent renditions as they are the only two types of rendition that can help us determine if there is some kind of improvement across interpreting tasks (increase in close renditions and decrease in divergent renditions). In other words, the other types of rendition, i.e. reduced, zero, expanded, summarized, substituted, and innovative, are left aside in this section. As a reminder, IN01 was used on November 3rd, IN02 on November 10th, IN03 on November 17th and IN04 on December 8th, 2016.

Table 24 – Overview of the percentages of close and divergent renditions

	IN01	**IN02**	**IN03**	**IN04**	**Total %**
Close rendition	50.68% (50/296)	58.42% (215/368)	52.13% (171/328)	52.29% (217/415)	53.38%
Divergent rendition	13.51% (40/296)	16.85% (62/368)	23.78% (78/328)	20.24% (84/415)	18.59%

As can be seen in Table 24, there is no clear trend and no marked improvement across time here. Although there is an increase in the percentage of close renditions from IN01 to IN02, the contrary cannot be said about the percentage of divergent renditions. In addition, close renditions drop to 52.12% in IN03 and poorly increase in IN04. In general, we can then say that the quality of the trainees' interpretations does not improve across time. However, we can take a closer look at the global features we investigated and see how delivery rates, lexical density scores and expansion rates can be linked to the trainees' proportion of close and divergent renditions.

Table 25 – Percentage of close and divergent renditions (of the five students under investigation)

		IN01	IN02	IN03	IN04	Total %
STU01	**Close rendition**	50.00% (37/74)	59.78% (55/92)	54.88% (45/82)	57.83% (48/83)	55.62%
	Divergent rendition	16.22% (12/74)	17.39% (16/92)	24.39% (20/82)	19.28% (16/83)	19.32%
STU02	**Close rendition**	52.70% (39/74)	69.56% (64/92)	54.88% (45/82)	59.04% (49/83)	59.04%
	Divergent rendition	10.81% (8/74)	16.67% (15/92)	19.51% (16/82)	15.66% (13/83)	15.57%
STU04	**Close rendition**	41.89% (31/74)	/	57.32% (47/82)	48.19% (40/83)	49.13%
	Divergent rendition	17.57% (13/74)	/	19.51% (16/82)	26.51% (22/83)	21.19%
STU06	**Close rendition**	/	46.74% (43/92)	41.46% (34/82)	43.37% (36/83)	43.86%
	Divergent rendition	/	18.48% (17/92)	31.71% (26/82)	24.10% (20/83)	24.76%
STU07	**Close rendition**	58.11% (43/74)	57.61% (53/92)	/	53.01% (44/83)	56.24%
	Divergent rendition	9.46% (7/74)	15.22% (14/92)	/	15.66% (13/83)	13.45%

It appears that when the student's delivery rate is much lower than the corresponding IN delivery rate, then the student correctly translates less than half of the occurrences. This is the case for STU06's three interpretations and STU04's interpretation of IN01. However, there are not enough examples to confirm that assumption and it would need deeper investigation. Furthermore, as said in Section 4.1.1, the delivery rate gradually increases and thereby increases the difficulty of the interpreting tasks to be performed by the students. IN03 and IN04 respectively have a delivery rate of 108.14 w/m and 109.9 w/m and the proportion of wrongly translated items is the most significant for these two source speeches. There is thus a direct link between a faster delivery rate and the proportion of divergent renditions.

As for lexical density, we can say that a high lexical density (LD) score seems to have a direct link with the proportion of wrongly translated items as IN03 has the highest LD score (57.83%) and the highest percentage of divergent renditions

(23.78%). However, the four IN lexical densities are between 52 and 58%, which means that all source speeches are rather dense, not just IN03. Another observation is the potential link between STU lexical density scores and STU divergent renditions. Indeed, there are four students (STU02, STU04, STU06, STU07) who support this idea as the IN for which their LD score is the lowest is also the IN for which their percentage of divergent renditions is the highest. Nevertheless, all STU lexical density scores are between 50% and 55% so the difference between LD scores of a same student is not significant at all. Finally, there is no potential direct link between close/divergent renditions and expansion rates.

In conclusion, there is no marked improvement in the quality of the trainees' interpretations considering their renditions of four linguistic items. In addition, some observations and assumptions have been made concerning potential links between some of the global features (lexical density and delivery rate) and close/divergent renditions. These observations though need deeper investigation.

It is very important to keep in mind that this is only one way to assess the quality of an interpretation and does not mean that they did not improve across time at all. It would be very interesting to investigate non-content related aspects of interpreting such as fluency (disfluency markers such as repetitions, self-corrections, hesitations or silent pauses), intonation, etc.

4.2.4 Case study: zero renditions

As I said throughout the present study, I think it is very interesting to focus on zero renditions, as omissions can be justified or unjustified, conscious or unconscious (Gile, 1995). For that reason, I decided to carry out a small-scale case study and to focus on zero renditions of two students, namely STU01 and STU02.

Viaggio (2002: 239) states that everything redundant, irrelevant, parasitic or incomprehensible should not be interpreted. Barik (1975) proposed a taxonomy including four types of omissions: skipping (of a lexical item), delay (omission of a larger unit of meaning due to lagging), comprehension (omission of a larger unit of meaning as a result of an inability to understand), and compounding (conjoining elements from different clauses). The last type of omission will be excluded from this analysis as it seems closer to a reduced rendition than a zero rendition.

Zero renditions on average account for 6.78% of STU01's and STU02's performances. What is interesting is that these two students do not omit the same type of linguistic item. STU01 seems to not interpret some **proper names** while STU02 appears to omit more **complex noun phrases**. Indeed, STU01 omitted 3 proper names: *Aldous Huxley* (IN02), *Snow* (IN02), and *Yogi Berra* (IN04). In the first case, STU01 omitted the proper name, which was the subject of the subclause "*as Aldous Huxley put it in this very room*", but also the rest of the sentence. In the second case, *Snow* was also the subject of the subclause "*what Snow called the two cultures*". STU01 omitted the first part and only translated "*the two cultures*". Finally, STU01 omitted *Yogi Berra* but succeeded in translating who he was "*a philosopher who was a very smart man despite being a Yankee*". The first and second errors have a more important impact on the output than the third one, as the message is still conveyed in the last example.

As for STU02, 20 complex noun phrases were omitted. If we take a closer look at the different sub-patterns, STU02 seems to omit more occurrences of the pattern "A + N" than of the other two. Indeed, 14 occurrences of "A + N" NPs were omitted such as *hidden agenda* (IN01), *classical Greek art* (IN03), *human genetics* (IN04) or *cerebral hemispheres* (IN02). In the first example, STU02 still conveyed the message as the original sentence was redundant "*it must have a hidden agenda, it must have another function*" and STU02 only interpreted the second part of the sentence. STU02 did not succeed in conveying the message in the second case and just omitted this unit of language, which altered the meaning of the original sentence.

> **Original sentence**: "… Especially they were influenced by Greek art classical Greek art and European Renaissance art where realism is strongly emphasized."
>
> **Interpreted sentence**: " … De même si on regarde la renai- les arts de re- de la Renaissance par exemple ehm ce ne sont pas les mêmes ehm standards esthétiques mais on vous dira aujourd'hui que ehm l'art n'a rien à voir avec le réalisme. »

Finally, it is interesting to see that some NPs were omitted by both students. For the first pattern (A + N), both students omitted *schizophrenic people* (IN01) and *Martian alphabet* (IN01). If we take a closer look at the first example, we see that the whole message was not conveyed in both interpretations.

> **Original sentence**: "… So you don't take the metaphor literally although schizophrenic people do but we don't and we're not going to do."
>
> **Interpreted sentence**: "… ehm ce genre de lien est donc totalement inespéré de la part de de gens qui ne qui ne qui ne le comprennent pas." (STU01)
>
> **Interpreted sentence** : « … Non on ne prend pas la métaphore au pied de la lettre." (STU02)

Investigating zero renditions is really interesting because we really need to take a close look at the student's output to know if the omission is justified or unjustified. We can conclude that the omissions presented in this small-scale study represent

the different types of omissions in Barik's taxonomy: there are some examples of skipping (*Snow*), of delay (*Aldous Huxley, schizophrenic people* – STU02) and also of comprehension (*schizophrenic people* – STU01). In some cases, the omission is completely justified and does not alter the message as the source text was redundant (*Martian alphabet*). However, in other cases, the omission is a bit more problematic as the original meaning is not completely translated (*Aldous Huxley*, *schizophrenic people*).

As this small-scale study only presents the linguistic categories with the most occurrences of zero renditions (**proper names** and **complex noun phrases**), it would be of a great interest to carry out further research on zero renditions for the other linguistic categories.

5 Conclusion

This thesis first aimed at finding out what the most challenging linguistic item was for interpreter trainees. It also focused on investigating students' interpreting quality to see if there was any marked improvement in the rendition of chosen linguistic items across time. **Complex noun phrases** turn out to be the most challenging linguistic item for interpreter trainees as only over half of all NPs were correctly translated. The "N + subclause" sub-pattern is the most difficult among the three investigated sub-patterns as it has a very low percentage of close renditions (45.28%) and a high percentage of divergent renditions (24.42%).

In order to answer the second research question (improvement in quality across time), I focused on five students as they carried out at least three interpreting tasks (STU01, STU02, STU04, ST06 and STU07) and on their proportion of close and divergent renditions across the interpreting tasks. Unfortunately, there is no clear trend and no marked improvement in the interpreting quality is to be observed across time. Global features can be linked to students' performances, but these observations would need more in-depth investigation. First, it seems that when the student's **delivery rate** is much lower than the corresponding IN delivery rate, the student correctly translates less than half of the occurrences. There also seems to have a direct link between a faster delivery rate and a higher proportion of divergent renditions. As for **lexical density**, STU lexical density scores and STU divergent renditions could be related. Indeed, there are four students who support this idea, as the IN for which their LD score is the lowest is also the IN for which their percentage of divergent renditions is the highest.

This study has of course its own limitations. It is first really important to keep in mind that this is only one way to assess the quality of an interpretation and does not mean that they did not improve across time at all. Results could also have been different if numbers were more complex for example. This study only focuses on different linguistic items, but my database also comprises the students'

disfluencies (repetitions, hesitations, self-corrections, silent pauses). These disfluency markers have not been investigated in the present study due to a lack of time, but it would be very interesting to link these to students' performances in a further research.

Speaking of future research, we have seen that **complex noun phrases** are the most challenging items for interpreter trainees and we investigated three types of NPs as they were the most frequent in the database: "A + N", "N + prep + N", and "N + subclause"'. However, it would be of a great interest to carry out a research on other types of NPs, such as "N + N" noun phrases, as this pattern is much more frequent in English than in French (Paillard, 2000: 49-51). It seems that this type of NPs can be very difficult for interpreter trainees as Lefer & De Clerck (2019) state that more than half of their "N + N" noun phrases had been simplified by professional interpreters. In addition, investigating other language pairs can also be very meaningful. It will then be possible to state if the source language, here English, has the same influence on interpreter trainees having different mother tongues (Italian or Spanish for example).

6 Appendices

6.1 Appendix 1: studies on interpreting based on authentic corpora (Setton, 2011)

Languages: EN English, ES Spanish, DE German, FI Finnish, FR French, HEB Hebrew, IT Italian, RU Russian, SW Swedish, TR Turkish, ZH (Standard) Chinese

Researcher	**Languages**	**Event**	**Mode**	**Subjects**	**Length**	**transcription published or available**	**Sound files availability**	**Analysis**
Oléron & Nanpon 1965	EN, FR, DE, ES	UNESCO impromptu (non-tech discussion)	SI	pros	~ 7 mins	Unknown	Unknown	Time lag, speed, fidelity
Déjean Le Feal 1978	FR>DE	Various speeches	SI	pros		77 pages, speakers and interpreters	Text on micro fiches, AIIC*	Recited vs impromptu input (école du sens)
Chernov 1979, forthcoming	EN, FR, ES, RU	UN 1968	SI	pros	"~40 hours"	'Parallel transcripts' extracts published	Probably N.A	Illustrate theory (redundancy & prediction)
	EN>RU, ES, FR	1978 UN satellite interpreting experiment	SI	pros		(ditto)	Probably no longer available	(ditto)
Lederer 1981	DE > FR	Railway Consortium and lab (2nd versions)	SI	2 pros	3 hrs taped (orig+ 2 interp)	63 mins original DE, FR; some extracts synchronised (interlinear)		Illustrate theory (école du sens)
Shlesinger 1989	HEB><EN	Courtroom testimony	SI	pros	4 hrs			
Donovan 1994	FR><EN	Extracts from 2 meetings 1986-8	SI	pros	~ 4 hrs	50 pp SI (2 events), fluent text speaker + interp, some extracts synchronised (interlinear)	No longer available	Fidelity examples (école du sens)
Pöchhacker 1994	EN><DE, FR >DE	Vienna small business conference *ICSB*	SI	pros	14 hrs (orig. + interpr.)	Available as vol. 3 of doctoral diss. from U. of Vienna library; parts published in Pöchhacker	Tapes available from author	Intertextuality, situational & delivery factors (speed, slips/shifts, hesitation

						94		EN><DE
Kalina 1998	EN><DE<>FR	*Bertell*:1989 public lecture (anti nuclear)	SI	6 students	70 mins	5 versions tiered. 5s per line.	Consult, loan at Heidlbrg	Choose examples, identify errors, strategies
	DE, FR, EN	*Würzburg* :1992 law symposium	SI	6 pros (2 per booth)		1-track audio (simulation: 2 track)	(ditto)	(ditto)
Setton 1997, 99	DE>EN	Extracts from Kalina *Würzburg* corpus	SI	1 pro + 2 in mock	14/30 min microanalysis	1-track audio (simulation: 2 track)	Available from author	Linguistic, cognitive-pragmatic analysis for process modelling
Wallmach 2000 & forthcoming	EN, Zulu, Afrikaans, Sepedi	Parliament speeches Gauteng (S. Africa) Provincial legislature	SI	16 pros	Pilot: 6-8 hrs EN, Afrikaans, Zulu	Pilot material transcribed (6-8 hrs)	110 hrs on tape (3 langs)	Norms/strategies vs. user expectations, effect of speed, technicality on performance, language-specific strategies
Cencini 2000	EN><IT	Television Interpreting Corpus (TIC),			36,000 wds	Not available for outside use		TEI Standard (computer query-able)
Fumagalli 1999-2000	EN>IT	Comparable EN (18) and IT source speeches;	CI			Not available for outside use		Features of 'interpretese' *MultiConcord-Parallel Concordancer*
Diriker 2001	EN><TR	Conference on Metaphysics and Politics (2 days)	SI	pros	150 pp transcripts	Available as Annex to Ph.D.	Bogazici Univ. Library	Interactional and sociological
ECIS Group (U. of Granada)	EN>DE,ES,FR; DE, FR>ES	European Parliament debates via EbS	SI	pros	43 speeches, 73 interpretations	Available from authors	Available from authors	Verbal and non-verbal minimizers and maximizers of quality, problem triggers, strategies
Vuorikoski 2004	mostly DE, EN ><F	European Parliament debates	SI	Ca. 70 pros	120 speeches, 65 analysed	Selected transcripts appended to	CD and website	Difficulty, quality in rhetorical

	I, SW					PhD		(political) speech genre
Beaton 2007		European Parliament debates via EbS	SI		7 hours			Text-discourse-ideology link through cohesive devices
Monacelli 2005, 2009	IT><EN	10 speeches from 4 events conferences	SI	10 pros	2 hrs origs + interpretation	Selected transcripts in PhD	CD ST, TT and synchronized (from audio cues)	Interactional politeness, 'face', through deixis, mood, transitivity
Straniero 2007	IT><EN	TV talk shows	SI and CI					Conversation Analysis
CIAIR (Nagoya University)	EN><JA	Monologue SI (simple lectures) and travel dialogues	SI and CI	SI: 4 pros per speech	182 hrs (1m wds)	Web access/interface	Not shared	Strategies, lag, etc. for machine SI research. Own software.
EPIC**: Bologna-Forli group	EN-IT-ES (9 sub-corpora)	European Parliament sessions	SI	Multiple pros	Open-ended; 280 hrs usable in archive; ca. 280, 000 wds trans cribed as of mid-2010	.Available for query on line (POS-tagged and indexed database)	As separate files, on request	As of 2010: directionality, lexis, dysfluencies; Ongoing (+ grad theses). Semi-automatic transcription technique
Meyer (2008) 'K6'	PT>DE	Lecture tour	SI, CI		6 hours, 35,000 wds transcribed (HIAT)	Web access, available at http://www.exmaralda.org/en_index.html		Proper names in CI, SI. Recording & transcription: EXMARaLDA
Meyer K2 corpus (forthcoming)	DE, TR, PT, ES	Doctor-patient dialogues in hospitals	Monolingual & interpreted		25 hrs recorded, 160,000 wds	Web access, available at http://www.exmaralda.org/en_index.html		
Shlesinger 2008, Shlesinger & Ordan 2010	EN	EPIC	SI, written translation					Features of 'interpretese' vs original speech and written translation
DIRSI (Bendazzoli 2010a)	EN><IT	International confs. (health) monologues	SI	5 pros (one EN-A)	2010: 20 hrs, 130,000 wds transcribed, tagged & indexed;	Accessible via dedicated online web interface.		Directionality

					text-to-sound and ST-TT aligned			
FOOTIE (Sandrelli forthcoming)	EN><IT	Football press confs: interactive Q&A	SI			In progress		Work begun in 2010
CorIt (Straniero and Falbo forthcoming)	multilingual	TV interpreting in Italy (longitudinal)	CI, SI			In progress		In progress

*The Research Committee of AIIC (International Association of Conference Interpreters) provides copies of microfiches of some theses to researchers on request.

**European Parliament Interpreting Corpus.

6.2 Appendix 2: Consent forms

Teacher

AGREEMENT FORM

LAST NAME	Bionda
First Name	Tania
Student Identification Number	/

I agree to take part in Thomas SIMON's data collection project.

I understand that my interpretations will be included in his database. They will not be shared with third-parties.

I understand that any data collected this year will be anonymized before being included in the database.

I hereby give permission for my interpretations to be included in the database and used for research purposes.

Date: 15th November 2016 Signature:

STUD1

AGREEMENT FORM

LAST NAME	VANDEN-BULCKE
First Name	Victor
Student Identification Number	8220-15-00

I agree to take part in Thomas SIMON's data collection project.

I understand that my interpretations will be included in his database. They will not be shared with third-parties.

I understand that any data collected this year will be anonymized before being included in the database.

I hereby give permission for my interpretations to be included in the database and used for research purposes.

Date: 17/11/2016

Signature:

SUO2

AGREEMENT FORM

LAST NAME	NEIRYNCK
First Name	Corentin
Student Identification Number	6865-15-00

I agree to take part in Thomas SIMON's data collection project.

I understand that my interpretations will be included in his database. They will not be shared with third-parties.

I understand that any data collected this year will be anonymized before being included in the database.

I hereby give permission for my interpretations to be included in the database and used for research purposes.

Date: 17.11.2016 Signature:

SN03

AGREEMENT FORM

LAST NAME	MOLDERS
First Name	CORENTIN
Student Identification Number	7224-15-00

I agree to take part in Thomas SIMON's data collection project.

I understand that my interpretations will be included in his database. They will not be shared with third-parties.

I understand that any data collected this year will be anonymized before being included in the database.

I hereby give permission for my interpretations to be included in the database and used for research purposes.

Date: 17/11/2016

Signature:

STU04

AGREEMENT FORM

LAST NAME	THEIS
First Name	Raphaël
Student Identification Number	7227-15-00

I agree to take part in Thomas SIMON's data collection project.

I understand that my interpretations will be included in his database. They will not be shared with third-parties.

I understand that any data collected this year will be anonymized before being included in the database.

I hereby give permission for my interpretations to be included in the database and used for research purposes.

Date: 17/11/2016 Signature:

STUDS

AGREEMENT FORM

LAST NAME	PELIKAN
First Name	Yann
Student Identification Number	31 88 07

I agree to take part in Thomas SIMON's data collection project.

I understand that my interpretations will be included in his database. They will not be shared with third-parties.

I understand that any data collected this year will be anonymized before being included in the database.

I hereby give permission for my interpretations to be included in the database and used for research purposes.

Date: 17.11.16

Signature:

STU06

AGREEMENT FORM

LAST NAME	BERTRON
First Name	Edouard
Student Identification Number	2654-16-00

I agree to take part in Thomas SIMON's data collection project.

I understand that my interpretations will be included in his database. They will not be shared with third-parties.

I understand that any data collected this year will be anonymized before being included in the database.

I hereby give permission for my interpretations to be included in the database and used for research purposes.

Date: 17.11.16

Signature:

STU08

AGREEMENT FORM

LAST NAME	DECLERCQ
First Name	Sophie
Student Identification Number	6780 - 15- 00

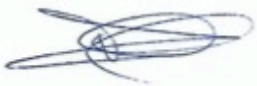
I agree to take part in Thomas SIMON's data collection project.

I understand that my interpretations will be included in his database. They will not be shared with third-parties.

I understand that any data collected this year will be anonymized before being included in the database.

I hereby give permission for my interpretations to be included in the database and used for research purposes.

Date: 17/11/2016 Signature:



6.3 Appendix 3: TreeTagger tagset description (English)

POS tag (EN)	Description	Lexical POS tag	Grammatical POS tag	Not relevant
CC	coordinating conjunction		**X**	
CD	cardinal number		**X**	
CDZ	possessive pronoun		**X**	
DT	determiner		**X**	
EX	existential *there*		**X**	
FW	foreign word	**X**		
IN	preposition, sub. conjunction		**X**	
IN/that	*that* as subordinator		**X**	
JJ	adjective	**X**		
JJR	adjective, comparative	**X**		
JJS	adjective, superlative	**X**		
LS	list marker			**X**
MD	modal		**X**	
NN	noun, singular or mass	**X**		
NNS	noun plural	**X**		
NNSZ	possessive noun, plural	**X**		
NNZ	possessive noun, singular or mass	**X**		
NP	proper noun, singular	**X**		
NPS	proper noun, plural	**X**		
NPSZ	possessive proper noun, plural	**X**		
NPZ	possessive noun, singular	**X**		
PDT	predeterminer (e.g. both the men)		**X**	
PP	personal pronoun		**X**	

PPZ	possessive pronoun		**X**	
RB	adverb	**X**		
RBR	adverb, comparative	**X**		
RBS	adverb, superlative	**X**		
RP	particle		**X**	
SENT	sentence-break punctuation			**X**
SYM	symbol			**X**
TO	infinitive “to”		**X**	
UH	interjection		**X**	
VB	verb *be*, base form (*aux.)	**X**		
VBD	verb *be*, past tense (*aux.)	**X**		
VBG	verb *be*, gerund/present participle	**X**		
VBN	verb *be*, past participle	**X**		
VBP	verb *be*, sing. present, non-3d (*aux.)	**X**		
VBZ	verb *be*, 3rd pers. sing. pres. (*aux.)	**X**		
VH	verb *have*, base form (*aux.)	**X**		
VHD	verb *have*, past tense	**X**		
VHG	verb *have*, gerund/present participle	**X**		
VHN	verb *have*, past participle	**X**		
VHP	verb *have*, sing. pres., non-3d (*aux.)	**X**		
VHZ	verb *have*, 3rd pers. sing. pres. (*aux.)	**X**		
VV	verb *have*, base form (*aux.)	**X**		
VVD	verb, past tense	**X**		
VVG	verb, gerund/present participle	**X**		

VVN	verb, past participle	**X**		
VVP	verb, singular present, non-3d	**X**		
VVZ	verb, 3rd person sing. present	**X**		
WDT	wh-determiner		**X**	
WP	wh-pronoun		**X**	
WPZ	possessive wh- pronoun		**X**	
WRB	wh-adverb		**X**	
Z	possessive ending		**X**	

Adapted from English TreeTagger PoS tagset with Sketch Engine modifications (n.d.)

6.4 Appendix 4: TreeTagger tagset description (French)

POS tag (FR)	Description	Lexical POS tag	Grammatical POS tag	Not relevant
ABR	abbreviation		**X**	
ADJ	adjective	**X**		
ADV	adverb	**X**		
DET:ART	article		**X**	
DET:POS	possessive pronoun (ma, ta, …)		**X**	
INT	interjection		**X**	
KON	conjunction		**X**	
NAM	proper name	**X**		
NOM	noun	**X**		
NUM	numeral		**X**	
PRO	pronoun		**X**	
PRO:DEM	demonstrative pronoun		**X**	
PRO:IND	indefinite pronoun		**X**	
PRO:PER	personal pronoun		**X**	
PRO:REL	relative pronoun		**X**	
PRP	preposition		**X**	
PRP:det	preposition plus article (au, du, aux, des)		**X**	
PUN	punctuation			**X**
SENT	sentence-break punctuation			**X**
VER:cond	verb conditional (*aux.)	**X**		
VER:futu	verb future (*aux.)	**X**		
VER:impf	verb imperfect (*aux.)	**X**		
VER:infi	verb infinitive (*aux.)	**X**		
VER:pper	verb past participle (*aux.)	**X**		
VER:ppre	verb present participle	**X**		

	(*aux.)			
VER:pres	verb present (*aux.)	**X**		
VER:simp	verb simple past	**X**		
VER:subi	verb subjunctive imperfect	**X**		
VER:subp	verb subjunctive present (*aux.)	**X**		

Some tags did not occur in the corpora used in this study. These tags are PUN:cit (punctuation citation), SYM (symbol) and VERB:impe (verb imperative).

7 Bibliography

Agar, M. (1996). *The Professional Stranger: An Informal Introduction to Ethnography*. Academic Press. American Heritage Dictionary. 2000. Houghton Mifflin.

Amato, A., Mack, G. (2011). Intepreting the Oscar Night on Italian TV: An Interpreters' Nightmare?. In *The Intepreters' Newsletter. 16, pp. 37-60.*

Alessandrini, M. (1990). Translating Numbers in Consecutive Interpretation: An Experimental Study. In *The Interpreters' Newsletter. 3*, pp. 77-80.

Ananiadou, S. (1994). A Methodology for Automatic Term Recognition. In *Proceeding COLING '94 Proceedings of the 15th Conference on Computational Linguistics*. 2, pp. 1034-1038.

Aston, G., Bernardini, S., Stewart, D. (2004). *Corpora and Language Learners*. Amsterdam / Philadelphia: John Benjamins.

Aston, G. (2016). How Corpora Can Help the Interpreter Walk the Tightrope. In *Corpus-based Approaches to Translation and Interpreting: from Theory to Applications.* G. Corpas Pastor, and M. Seghiri (eds.). Frankfurt: Peter Lang. pp. 219-238.

Baker, M. (1992). *In Other Words: a Coursebook on Translation*. London and New York: Routledge.

Baker, M. (1993). Corpus Linguistics and Translation Studies: Implications and Applications. In *Text and Technology: In Honour of John Sinclair*. Baker, M., Francis, G. & E. Tognini-Bonelli (eds.). pp. 233-252. Amsterdam / Philadelphia: John Benjamins.

Baker, M. (1995). Corpora in Translation Studies. An Overview and Suggestions for Future Research. In *Target. 7*(2), pp. 223-243.

Bakti, M. (2009). Speech Disfluencies in Simultaneous Interpretation. In *Selected Papers of the CETRA Research Seminar in Translation Studies 2008*. D. De Crom (ed.).

Baldwin, T. (2010). Multiword Expressions. In *Handbook of Natural Language Processing*. N. Indurkhya & F. J. Demareau (eds.). pp. 267-292.

Bale, R. (2013). Undergraduate Consecutive Interpreting and Lexical Knowledge: The role of spoken corpora. In *The Interpreter and Translator Trainer*. 7(1), pp. 27-50.

Barik, H. C. (1975). Simultaneous Interpretation: Qualitatitve and Linguistic Data. In *Language and Speech*. 18, pp. 272-297.

Bartlomiejczyk, M. (2006). Strategies of Simultaneous Interpreting and Directionality. In *Interpreting.* 8(2), pp. 149–74.

Beaton, M. (2007). Interpreted Ideologies in Institutional Discourse: The Case of the European Parliament. In *The Translator* 13(2), pp. 271-296.

Bendazzoli, C. (2010). *Corpora e Interpretazione Simultanea*. Bologna: Asterisco.

Bendazzoli, C. (2010). The European Parliament as a Source of Material for Research into Simultaneous Interpreting: Advantages and Limitations. In *Translationswissenschaft – Stand und Perspektiven. Innsbrucker Ringvorlesungenzur Translationswissenschaft VI (Forum Translationswissenschaft, Bd. 12)*. N.L. Zybatow (ed.). pp. 51-68. Frankfurt a. M.: Peter Lang.

Bendazzoli, C. (2018). Corpus-based Interpreting Studies: Past, Present and Future Developments of a (Wired) Cottage Industry. In *Making Way in Corpus-based Interpreting Studies*. M. Russo et al. (eds). pp. 1-19. Singapore: Springer.

Bendazzoli, C., Sandrelli, A. (2005). An approach to Corpus-based Interpreting Studies: Developing EPIC (European Parliament Interpreting Corpus). In *MuTra 2005 – Challenges of Multidimensional Translation: Conference proceedings.*

Bendazzoli, C., Sandrelli, A. (2006). Tagging a Corpus of Interpreted Speeches: the European Parliament Interpreting Corpus (EPIC). In *Proceedings of the LREC 2006 Conference*. *Genova. Magazzini del Cotone 24-26 May 2006.*

Bendazzoli, C., Sandrelli, A. (2009). Corpus-based Interpreting Studies: Early Work and Future Prospects. In *Revista Tradumàtica. L'aplicaciò dels corpus lingüístics a la traducció 7.* Accessed 15 April 2019.

Bernardini, S. Ferraresi, A. & Miličević, M. (2016). From EPIC to EPTIC: Exploring Simplification in Interpreting and Translation from an Intramodal Perspective. In *Target*. 28(1), pp. 61-86.

Biber, D., Johansson S., Leech G., Conrad S. & Finegan E. (1999). *Longman Grammar of Spoken and Written English*. Harlow, Longman.

Braun, S. & Clarici, A. (1996). Inaccuracy for Numerals in Simultaneous Interpretation: Neurolinguistic and Neuropsychological Perspectives. In *The Interpreters' Newsletter.* 7, pp. 85–102.

Cencini, M. (2002). On the Importance of an Encoding Standard for Corpus-based Interpreting Studies: Extending the TEI scheme. In *CULT2 K. Special Issue of InTRAlinea. S.* Bernardini, F. Zanettin (eds.)

Chernov, G. V. (1978) *Teoriya i praktika sinkronnogo perevoda* (Theory and Practice of Simultaneous Interpretation). Moscow: Mezhdunarodniye otnosheniya.

Chernov, G. V. (1987). *Fundamentals of Simultaneous Interpreting*. Moscow, Vysshaya Shkola.

Chernov, G. V. (2004). *Inference and Anticipation in Simultaneous Interpreting. A probability-prediction model*. Amsterdam/Philadelphia: John Benjamins.

Chmiel, A., Mazur, I. (2013). Eye Tracking Sight Translation Performed by Trainee Interpreters. In *Tracks and treks in translation studies*. C. Way, S. Vandepitte, R. Meylaerts & M. Bartłomiejczyk (eds.). Amsterdam: John Benjamins. pp. 189–205.

Collard, C., Defrancq, B. (2019a). Disfluencies in Simultaneous Interpreting: a Corpus-based Study with Special Reference to Sex. In *New empirical perspectives on translation and interpreting*. Presented at the Translation in Transition 3.

Collard, C., Defrancq, B. (2019b). Corpus Study on the Rendition of Numbers in Simultaneous Interpreting, with Special Reference to Sex. In *Interpreting.*

Communicate (1999). AIIC Webzine n°1. *Not available anymore*. Seen in Li (2010).

Concise Oxford Dictionary. [CD-ROM]. Oxford: Oxford University Press, 1996.

Dal Fovo, E. (2011). Through the CorIT Looking-glass and What MA Students Found There. In *The Interpreters' Newsletter.* 16, pp. 1-20.

Dastjerdi, H. (2011). Translation of Idioms: a Hard Task for the Translator. In *Theory and Practice in Language Studies*. 1(7), pp. 879-883.

David, C. (2008). *A Dictionary of Linguistics and Phonetics*. (6th edition). Blackwell Publishing.

Dayter, D. (2016). Corpus-based Approach to Simultaneous interpretation at the United Nations: Multidimensional analysis of variation. *Paper presented at the EST Congress 2016, Aarhus, Denmark, 15-17 September 2016.*

Dejean Le Féal, K. (1982). Why impromptu speech is easy to understand. In *Impromptu Speech*. N. E. Enkvist (ed.). pp. 221-230.

Dembry, C., Love, R. (2015). Collecting the Spoken BNC2014 – Overview of Methodology*. Paper presented at the Corpus Linguistics 2015 Conference, Lancaster University, UK, 21-24 July 2015*.

Desmet, B, Vandierendonck M., & Defrancq B. (2018). Simultaneous Interpretation of Numbers and the Impact of Technological Support. In *Interpreting and technology*. Claudio Fantinuoli (ed.). pp. 13-27. Berlin: Language Science Press.

Ding Z., Ryu K., Matsubara, S. & Yoshikawa, M. (2005). Interpreting Unit Segmentation of Conversational Speech in Simultaneous Interpretation Corpus. In *Proceedings, COCOSDA 2005 (International Committee for the Co-ordination and Standardization of Speech Databases and Assessment Techniques)*, December 2-3, Bali, Indonesia, pp. 148-152.

Diriker, E. (2001). *Contextualising Simultaneous Interpreting: Interpreters in the Ivory Tower?* Doctoral dissertation. Bogazici University, Turkey.

Diriker, E. (2004). *De-/Re-contextualizing conference interpreting: Interpreters in the ivory tower?.* Amsterdam: John Benjamins.

Donato V. (2003). Strategies Adopted by Student Interpreters in SI: a Comparison between the English-Italian and the German-Italian Language-pairs. In *The Interpreters' Newsletter.*12, pp. 101-134.

Fernandes, L. (2006). Corpora in Translation Studies: Revisiting Baker's typology. In *Fragmentos* 30, pp. 87-95.

Fumagalli, D. (1999-2000). *Alla ricerca dell'interpretese. Uno studio sull'interpretazione consecutiva attraverso la corpus linguistics*. Unpublished dissertation. Advanced School for Translators and Interpreters (SSLMIT), University of Trieste.

Gardner, D. & Davies, M. (2007). Pointing out Frequent Phrasal Verbs: A Corpus-based Analysis. In *Tesol Quarterly.* 41(2), pp. 339–359.

Galli, C. (1990). Simultaneous Interpretations in Medical Conferences: a Case Study. In *Aspects of Applied and Experimental Research on Conference Interpretation*. L. Gran & C. Taylor (eds.). pp. 61-82. Udine: Campanotto Editor.

Gerver D. (1971). The Effects of Source Language Presentation Rate on the Performance of Simultaneous Conference Interpreters. In *Proceedings of the 2nd Louisville Conference on Rate and/or Frequency Controlled Speech*. E. Foulke (ed.). University of Louisville, pp. 162-184.

Gerver, D. (1976). Empirical Studies of Simultaneous Interpretation: a Review and a Model. In *Translation: Applications and Research*. R.W. Brislin (ed.). pp. 165-207. New York: Gardner.

Ghiselli, S. (2018). The Translation Challenges of Premodified Noun phrases in Simultaneous Interpreting from English into Italian. In *InTRAlinea Special Issue: New findings in corpus-based interpreting studies*.

Gile, D. (1992). Predictable Sentence Endings in Japanese and Conference Interpretation. In *The Interpreter's Newsletter*, Special Issue n°1. pp. 12-23.

Gile, D. (1995). *Basic Concepts and Models for Interpreter and Translator Training*. Amsterdam: Benjamins.

Gile, D. (1999). Testing the Effort Models' Tightrope Hypothesis in Simultaneous Interpreting – a Contribution. In *Hermes*. 23, pp. 153-172.

Gile, D. (2008). Local Cognitive Load in Simultaneous Interpreting and its Implications for Empirical Research. *Forum*. 6, pp. 59–77.

Goldman-Eisler, F. (1972). Segmentation of Input in Simultaneous Translation. In *Journal of Psycholinguistic Research.* 1(2), pp. 127-140.

Gottlieb, H. (1997). You Got the Picture? On the Polysemiotics of Subtitling Wordplay. In *Essays on Punning and Translation*. pp. 207-232. Manchester & Namur: St Jerome.

Halliday, M. A. K. (1985). *Spoken and Written Language*. Geelong Vict.: Deakin University. Deakin University (ed.).

Halverson, S. (1998). Translation Studies and Representative Corpora: Establishing Links between Translation Corpora, Theoretical/Descriptive Categories and a Conception of the Object of Study. In *Meta*, 43(4).

Hanaoka, O. (2002). Proper Names in the Media: Problems for Translators/Interpreters and L2 Learners. In *Interpretation Studies*. 2, pp. 28-42.

Harvey, M. (2000). *A Beginner's Course in Legal Translation: the Case of Culture-bound Terms*. Université Lumière Lyon 2, France.

Herbert J. (1952). *The Interpreter's Handbook. How to Become a Conference Interpreter*. Genève, Librairie de l'Université George.

Helmut, S. (1994). Probabilistic Part-of-Speech Tagging Using Decision Trees. In *Proceedings of International Conference on New Methods in Language Processing*, Manchester, UK.

Iacovoni, A. (2010). *Compression in Simultaneous Interpreting: Fact or Myth? A Corpus-based Study*. Retrieved 1 August 2019 from http://iacovini.files.wordpress.com/2010/ 8/tesi.pdf.

Johansson, S. (2007). Seeing through Multilingual Corpora: On the Use of Corpora in Contrastive Studies. Studies in Corpus Linguistics, 26. Amsterdam: John Benjamins.

Johansson, V. (2008). Lexical Diversity and Lexical Density in Speech and Writing: a Developmental Perspective. Lund University, Dept. Of Linguistics and Phonetics. *Working papers*. 53, pp. 61-79.

Jurafsky, D., Martin, J. H. (2004). Word Classes and Part-of-Speech Tagging. (Revised version) In *Speech and Language Processing: an Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*.

Kader, S. & Seubert, S. (2015). Anticipation, Segmentation...Stalling? How to Teach Interpreting Strategies. In *To Know How To Suggest... Approaches to Teaching Conference Interpreting.* D. Andres, M. Behr (eds). pp. 25–44. Berlin: Frank & Timme,

Kajzer-Wietrzny, M. (2012). *Interpreting Universals and Interpreting Style*. PhD dissertation. Adam Mickiewicz University.

Kajzer-Wietrzny, M. (2013). Idiosyncratic Features of Interpreting Style. In *New Voices in Translation Studies.* 9, pp. 38-52.

Kajzer-Wietrzny, M. (2018). Intepretese vs. Non-native Language Use: the Case of Optional *That*. In *Making way in corpus-based interpreting studies*. M. Russo et al. (eds). pp. 97-113. Singapore: Springer.

Kalina S. (1992). Discourse Processing and Interpreting Strategies: an Approach to the Teaching of Interpreting. In *Teaching Translation and Interpreting. Training, Talent and Experience*. C. Dollerup and A. Lindegaard (eds.). pp. 251-257*.* Amsterdam-Philadelphia, John Benjamins.

Kalina, S. (1998). *Strategische Prozesse beim Dolmetschen.* Tübingen: Gunter Narr.

Kennedy, G. (1998). *An Introduction to Corpus Linguistics*. London/New York: Longman.

Khan, S. (2016). The Distinction between Term and Word: a Translator and Interpreter Problem and the Role of Teaching Terminology. In *Procedia – Social and Behavioral Sciences*. 232, pp. 696-704.

Kilgariff, A., Rychly, P., Tugwell, D. (2004). The Sketch Engine. In *Proceedings Eleventh EURALEX International Congress*. Lorient, France.

Korpal, P. & Stachowiak, K. (2018). The Whole Picture: Processing of Numbers and their Context in Simultaneous Interpreting. In *Poznań Studies in Contemporary Linguistics.* 54(3), pp. 335-354.

Kruger, A. (2004). Corpus-based Translation Studies: Research and Applications. In *Special Issue of Language Matters. Studies in the Languages of Africa.* 35(1).

Laplace, C. (1994). *Théorie du langage et théorie de la traduction. Les concepts-clés de trois auteurs : Kade (Leipzig), Coseriu (Tübingen), Seleskovitch (Paris).* Paris: Didier-Erudition.

Laviosa, S. (1996). The Corpus-based Approach: a New Paradigm in Translation Studies. In *Meta.* 43(4), pp. 474-479.

Laviosa, S. (1998) Core Patterns of Lexical Use in a Comparable Corpus of English Narrative Prose. In *Meta*. (43)4, pp. 557-570.

Laviosa, S. (2000). Simplification in the Language of Translation: Before and After the Advent of Corpora. In *Athanor.* 10(3).

Laviosa, S. (2002). *Corpus-based Translation Studies: Theory, Findings, Application*. Amsterdam: Rodopi.

Laviosa, S. (2011). Corpus-based Translation Studies: Where Does It Come From? Where Is It Going? In *Corpus-based Translation Studies. Research and applications.* A. Kruger, K. Wallmach, and J. Munday (eds.). pp. 13-32. London: Continuum.

Laviosa, S. (2012). Corpora and Translation Studies. In *Corpus applications in Applied Linguistics*. K. Hyland, C. Meng Huat, and M. Handford (eds.). pp. 67-83. London: Bloomsbury.

Laviosa, S. (2015). Corpora and Holistic Cultural Translation. In *Corpus-based Translation and Interpreting Studies: From description to application.* M.T. Sánchez Nieto (ed.). pp. 31-51. Berlin: Frank & Timme.

Lederer, M. (1981). *La traduction simultanée*. Paris: Minard Lettres Modernes.

Leech, G. (1992). 100 Million Words of English: The British National Corpus (BNC). In *Language Research*. 28(1), pp. 1-13.

Lefer, M-A., De Clerck, M. (2019). *L'apport des corpus intermodaux en lexicologie contrastive : Étude comparative de la traduction écrire et de l'interprétation simultanée des séquences des noms*. Unpublished article.

Li, C. (2010). Coping Strategies for Fast Delivery in Simultaneous Interpretation. In *The Journal of Specialised Translation*. 13, pp. 19-25.

Liontou K. (2011). Strategies in German-to-Greek Simultaneous Interpreting: A Corpus-Based Approach. In *Gramma: Journal of Theory & Criticism.* 19, pp. 37-56.

Mackintosh, J. (1995). A Review of Conference Interpretation: Practice and Training. In *Target.* 7(1), pp. 119-133.

Malmkjær, K. (1998). Love thy Neighbour: Will Parallel Corpora Endear Linguists to Translators?. In *Meta*. 43(4), pp. 534-541.

Mazza, C. (2001). Numbers in Simultaneous Interpretation. In *The Interpreters' Newsletter.* 11, pp. 87-104.

McEnery, A., Xiao, R., Tono, Y. (2006). *Corpus-based Language Studies: An Advanced Resource Book*. C. Candlin, R. Carter (eds.). London: Routledge.

Mead, P. (2015). Numbers. In *Routledge Encyclopedia of Interpreting Studies*. Abingdon/New York: Routledge. pp. 286-288.

Metzger, M., Roy, C. (2011). The First Three Years of a Three-year Grant. When a Research Plan Doesn't Go As Planned. In *Advances in interpreting research: Inquiry in action.* B. Nicodemus, L. Swabey (eds.). pp. 59-84. Amsterdam: John Benjamins.

Monacelli, C. (2009). *Self-preservation in Simultaneous Interpreting. Surviving the role.* Amsterdam: John Benjamins.

Monti, C, Bendazzoli, C., Sandrelli, A., Russo, M. (2005). Studying Directionality in Simultaneous Interpreting Through an Electronic Corpus: EPIC (European Parliament Interpreting Corpus). In *Meta.* 50(4).

Moser, B. (1978). Simultaneous Interpretation: a Hypothetical Model and its Practical Application. In *Language Interpretation and Communication*. D. Gerver & H. Wallace Sinaiko (eds). Nato Conference Series, Series III: Human Factors. New York and London: Plenum Press. pp. 353-368.

Moser-Mercer, B. (1994). Paradigms Gained or the Art of Productive disagreement. In *Bridging the Gap: Empirical Research in Simultaneous Interpretation*. S. Lambert & B. Moser- Mercer (eds.). Amsterdam-Philadelphia, Benjamins. pp. 17-23.

Newmark, Peter (1988). *A Textbook of Translation*. Prentice Hall International: Hertfordshire.

Niemants, N. (2012). The Transcription of Interpreting Data. In *Interpreting*. 14(2), pp. 165-191.

Ohno, T., Tohyama, H., Matsubara, S. (2008). Construction and Analysis of Word-level Time-aligned Simultaneous Interpretation Corpus. In *Proceedings of European Language Resources Association (ELRA) Conference, Marrakesh, May 28-30 2008* (LREC 2008).

Oléron, P., Nanpon, H. (1965). Recherches sur la traduction simultanée. In *Journal de psychologie normale et pathologique.* 62, pp. 73-94.

Olohan, M. (2004). *Introducing Corpora in Translation Studies*. London/New York: Routledge.

Oluchukwu, A. (2015). Translation Challenges with Phrasal Verbs. Nnamdi Azikiwe University, Awka, Nigeria.

Paillard, M. (2000). *Lexicologie contrastive anglais-français: Formation des mots et construction du sens*. Paris/Gap, Ophrys.

Paneth, E. (1957). *An Investigation into Conference Interpreting (with Special Reference to the Training of Interpreters).* Master's thesis, University of London.

Pearson, J. (1998). *Terms in Context*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Pinochi, D. (2009). Simultaneous Interpretation of Numbers: Comparing German and English to Italian. An Experimental Study. In *The Interpreters' Newsletter.* 14, pp. 33–57.

Plevoets, K., Defrancq, B. (2018). The Cognitive Load of Interpreters in the European Parliament, a Corpus-based Study of Predictors for the Disfluency *uh(m)*. In *Interpreting*. 20(1), pp. 1-28.

Plevoets, K., Defrancq, B. (2018). Lexis or Parsing? A Corpus-based Study of Syntactic Complexity and its Effect on Disfluencies in Interpreting. Presented at the UCCTS 5.

Pöchhacker, F. (1994). *Simultandolmetschen als komplexes Handeln.* Tübingen: Günter Narr.

Pöchhacker, F. (1999). Teaching Practices in Simultaneous Interpreting. In *The Interpreters' Newsletter*. 9, pp. 157-176

Pradas Macías, E. (2009). Identificación del patrón pausístico para la medición de la calidad en interpretación simultánea. In *Translatione via facienda*, Festshrift für Christiane Nord zum 65. pp. 235-252. Frankfurt: Peter Lang.

Quini, H. (1993). English Broadcast Interpreting in Japan. In *Tsuyaku-riron Kenkyu [Interpreting Research: The Journal of the Interpreting Research Association of Japan].* 3(1), pp. 38-45.

Rodero, E. (2012). A Comparative Analysis of Speech Rate and Perception in Radio Bulletins. In *Text and Talk*. 32(3), pp. 391-411.

Riccardi A. (2005). On the Evolution of Interpreting Strategies in Simultaneous Interpreting. In *Meta*. 50(2), pp. 753-767.

Rose, R. (2017). Proceedings of DiSS 2017, 18–19 August 2017, Royal Institute of Technology, Stockholm Sweden. R. Eklund, R. Rose (eds.). TMH-QPSR, 58(1).

Rumelhart, D. E. (1980). Schemata: The Building Blocks of Cognition. In *Theoretical issues in reading comprehension*. R. Spiro, B. Bruce, W. Brewer (eds.). Erlbaum Associates. pp. 33-58.

Sandrelli, A., Bendazzoli, C. (2005). Lexical Patterns in Simultaneous Interpreting: a Preliminary Investigation on EPIC (European Parliament Interpreting Corpus). In *Proceedings from the Corpus Linguistics conference series.* Corpus Linguistics 2005.

Sandrelli, A., Russo, M., Bendazzoli, C. (2006). Looking for Lexical Patterns in a Trilingual Corpus of Source and Interpreted Speeches: Extended Analysis of EPIC (European Parliament Interpreting Corpus). In *Forum.* 4(1), pp. 221-254.

Sandrelli, A. (2010). Corpus-based Interpreting Studies and Interpreting Training: A modern proposal. In *Translationswissenschaft – Stand und Perspektiven. Innsbrucker Ringvorlesungenzur Translationswissenschaft VI (Forum Translationswissenschaft, Bd. 12)*. L.N. Zybatow (ed.). pp. 69-90. Frankfurt a. M.: Peter Lang.

Sarcevic, S. (1985). Translation of Culture-bound Terms in Law. In *Multilingual.* 4(3), pp. 127-133.

Scardulla, C. (2016). ELF Interpreting at the European Union: A Corpus-based Study. *Paper presented at the EST Congress 2016, Aarhus, Denmark, 15-17 September 2016.*

Sdobnikov V. (2017). Types of Decompression in Simultaneous Interpreting. In *Journal of Siberian Federal University. Humanities and Social Sciences*. 3(10), pp. 409-417.

Seleskovitch, D. (1962). L'interprétation de conférence. In *Babel.* 8(1), pp. 13-18.

Seleskovitch, D. (1975). *Langage, langues et mémoire: études de la prise de notes en interprétation consécutive.* Paris: Minard Lettres Modernes.

Setton, R. (1997). *A Pragmatic Theory of Simultaneous Interpretation*. Unpublished doctoral dissertation. The Chinese University of Hong Kong, Graduate Division of English/Applied Linguistics.

Setton R. (1999). *Simultaneous Interpretation. A Cognitive-Pragmatic Analysis*. Amsterdam-Philadelphia, John Benjamins.

Setton, R. (2011). Corpus-based Interpreting Studies (CIS): Overview and Prospects. In *Corpus-based translation studies. Research and applications*. A. Kruger, K. Wallmach and J. Munday (eds.). pp. 33-75. London: Continuum.

Shlesinger, M. (1998). Corpus-Based Interpreting Studies as an offshoot of Corpus-Based Translation Studies. In *Meta.* 43(4), pp. 486-493.

Shlesinger, M., Ordan, N. (2012). More Spoken or More Translated? Exploring a Known Unknown of Simultaneous Interpreting. In *Target.* 24(1), pp. 43-60.

Shojaei, A. (2012). Translation of Idioms and Fixed expressions: Strategies and Difficulties. In *Theory and practice in language studies*. 2(6), pp. 1220-1229.

Spinolo, N. (2018). Studying Figurative Language in Simultaneous Interpreting: The IMITES (Interpretación de la Metáfora Entre ITaliano y ESpañol) Corpus. In *Making way in corpus-based interpreting studies*. pp. 133-155. Singapore: Springer.

Straniero Sergio, F. (2007). *Talkshow Interpreting. La mediazione linguistica nella conversazione spettacolo*. Trieste: Edizioni Universitarie Trieste.

Timarová, S., Čeňková, I., Meylaerts, R., Hertog, E., Szmalec, A., & Duyck, W. (2014). Simultaneous Interpreting and Working Memory Executive Control. In *Interpreting.* 16*(2*), pp. 139– 168.

To, V., Fan, S. & Thomas, DP. (2013a). Lexical Density and Readability: a Case Study of English Textbooks. In *The International Journal of Language, Society and Culture*. 37(7), pp. 61-71.

Tohyama, H., Matsubara, S. (2006a). Collection of Simultaneous Interpreting Patterns by Using Bilingual Spoken Monologue Corpus. In *Proceedings of 5th International Conference on Language Resources and Evaluation, May 2006 (LREC-2006),* 2564-2569.

Tohyama, H., Matsubara, S. (2006b). Development of Web-based Teaching Material for Simultaneous Interpreting Learners using Bilingual Speech Corpus. In

Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications, June 2006 (ED-MEDIA-2006). pp. 2906-2911.

Ure, J. (1971). Lexical Density and Register Differentiation. In *Applications of linguistics. Selected papers of the Second International Congress of Applied Linguistics. Cambridge*. G.E. Perren & J. L. M. Trim (eds). pp. 443-452. Cambridge: Cambridge University Press.

Van Dam, I. (1989). Strategies of Simultaneous Interpretation. In *The Theoretical and Practical Aspects of Teaching Conference Interpretation: First International Symposium on Conference Interpreting at the University of Trieste*. L. Gran and J. Dodds (eds.). pp. 167–76. Udine, Campanotto.

Viaggio, S. (2002). The Quest for Optimal Relevance: the Need to Equip Students with a Pragmatic Compass. In *Interpreting in the 21st Century: Challenges and Opportunities*. G. Garzone, M. Viezzi (eds.). pp. 229-244. Amsterdam/Philadelphia.

Vuorikoski, A.R. (2004). *A Voice of its Citizens or a Modern Tower of Babel?.* Tampere: Tampere University Press.

Weston, M. (1991). *An English Reader's Guide to the French Legal System*. New York/Oxford, Berg.

Zanettin, F. (1998). Bilingual Comparable Corpora and the Training of Translators. In *Meta.* 43(4).

Zanettin, F., Bernardini, S. & Stewart, D. (2003). *Corpora in Translator Education*. Manchester / Northampton: St Jerome.

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**