To Bidouche

Promotor Prof. dr. Bart Defrancq
Vakgroep Vertalen, tolken en communicatie

Decaan Prof. dr. Gita Deneckere
Rector Prof. dr. Rik Van de Walle

Nederlandse vertaling:
Corpusgebasserde onderzoek over het simultaan tolken

FACULTEIT LETTEREN
EN WIJSBEGEERTE

A corpus-based study of simultaneous interpreting with special reference to sex

Camille Collard

Proefschrift voorgelegd tot het behalen van de graad van Doctor in de Wijsbegeerte
2018

UNIVERSITEIT
GENT

Acknowledgements

> « Perfectionism is nothing more
> than making the most of your
> skills, tools and time.”
>
> Joseph Collard

I’ve always been a strong defendant of concision, be it in my personal life or in my career as an interpreter (says the person who is about to hand out a 200-page document). Unfortunately for me, the list of people I’m grateful for is as long as a that of a Thanksgiving dinner table.

First and foremost, I would like to thank my supervisor Bart Defrancq for convincing me to embark on this journey, whose most enriching part has probably been our long exchanges on this fascinating profession of ours. I have a profound admiration for his dedication, his expertise and his teaching talent. Most importantly, I would like to thank Bart for supporting, guiding and advising me in a particularly respectful manner on this bumpy road.

Secondly, I would like to thank my father for the support he provided. This dissertation would not be the same without his incredible IT knowledge. I would also like to thank my mother for her interest in my work, and for always being supportive. In general, I would like to thank my family and friends for being there when I needed help, and for being distant when I needed focus. You made this journey incredibly bearable.

Knowing when to ask for help has never been my strong suit. That is why I’m particularly thankful for my friend and family who have spontaneously offered their help, saving me for the soul-crushing exercise of having to desperately cry for help. While I could have thanked them for so many other reasons, the support they gave me during the last days of my thesis award them the noble title of heroes.

Lauran, Alizée, Selim and Girlou, this thesis would not have seen the daylight (and probably neither would I) without you. THANK YOU.

I would like to thank Mieke Van Herreweghe, Renata Enghels, Koen Plevoets and Rob Hartsuiker for their valuable feedback and guidance throughout this journey. I would like to thank my colleague Cédric Magnifico for encouraging me to seize this opportunity. Our exchanges on our respective theses have made me feel understood and supported.

I would like to thank the other inspiring researchers I met along the road during exciting conferences. Most specifically I would like to thank Heike, co-author of one of the articles included in this thesis. Writing an article with four hands is not an easy task but she made it incredibly enjoyable. I would like to thank Alexis for enlightening me on the brains of interpreters. Thank you also to all those who have helped me during the last three years and mostly during the last days: Amy, Colette, Mehdi, Julia, Hanne, Paul and many more whom I too exhausted to remember at this point.

I would like to thank my friends and colleagues who had to witness this journey against their will. I would also like to thank my almost lifelong mentors, Chris and Julia for always being there. A special thanks to my best friend Selim for showing up for coffee and offering to cook when I was under time pressure. Thank you also to my flatmates and friends, Pierre and Olivier, for being so quiet and patient. Finally, thank you to my unbearable partner in crime, Gaspar, for teaching me to only rely on myself in times of crisis.

Abstract

Simultaneous interpreting boasts a long tradition of experimental research, which has allowed for the cognitive processes involved in interpreting to be more clearly specified: speech comprehension and production, memory, attention/resource allocation and coordination. However, experiments' small sample sizes and lack of ecological validity mean that findings are not always conclusive. This article-based doctoral dissertation presents the results of four quantitative corpus-based studies on interpreters at the European Parliament. In total, 220 interpretations in seven language pairs have been explored. While dozens of input and output variables were analysed, the main control variable for this research project is the interpreter's sex. Indeed, studies on individual cognitive tasks suggest that women score higher than men in several verbal tasks involved in the cognitive processes of simultaneous interpreting. These processes have been operationalised into four different research topics: the length of the Ear-Voice Span, the production of disfluencies, the rendition of numbers and the use of extrapositions by interpreters working into Dutch and German. For each topic, the influence of sex has been analysed, alongside several other input and output variables, such as source and target languages, delivery rates and frequency and duration of silent pauses. Results show that sex is not a significant predictor of the Ear-Voice Span, the use of extrapositions and the rendition of numbers. However, male interpreters appear to produce more disfluencies than female interpreters, especially filled pauses and lengthenings.

This research project has also allowed to confirm or refute findings from previous studies, as well as yielding some interesting new insights into the ways interpreters work in a real-life environment. Results of the first corpus study show that the length of EVS is in line with previous findings, with a median value of three seconds. Our data also indicate that the EVS increases when the input rate increases. Dutch produces longer EVS, which is in line with findings indicating that SOV source languages require a longer EVS. The EVS has also been found to increase the production of most disfluencies. As expected in the second corpus study, the frequency of errors and omissions in the rendition of numbers was much lower in our data compared to experimental data, given that interpreters often receive help from their colleagues or the written speech. Our findings confirm that omissions are the most frequent category after complete rendition and that omissions increase if the input delivery rate increases. However, results find no significant influence of the nature and complexity of numbers, as well as source and target languages on the rendition of numbers. The third corpus study shows that interpreters produce more disfluencies than source speakers, which tends to confirm that interpreting is cognitively more demanding than spontaneous speech. The frequency of disfluencies produced by the interpreters appears to increase when the source speaker's delivery rate increases, which is in line with literature indicating that the source

speaker’s delivery increases the cognitive load. The last corpus study shows that interpreters into Dutch and German tend to shorten their middle field more than original speakers and use extraposition partly because they imitate the clause word order of the source speech (in this case French). These findings confirm that interpreters try to save memory capacity. Moreover, German-speaking source speakers have shorter middle field, compared to their Dutch-speaking counterparts, confirming a tendency documented in the literature.

Thanks to the methodologies applied in order to transcribe, time-tag and annotate the present corpus, as well as to analyse and quantify the results, we hope to inspire other researchers to use corpus data to observe cognitive processes in interpreting. More specifically we encourage researchers to include sex as a predictor variable in their experimental designs in order to confirm or refute our findings.

Samenvatting

Het simultaan tolkproces kan bogen op een lange traditie van experimenteel onderzoek, waardoor de cognitieve processen van het tolken duidelijker werden gespecificeerd: spraakverstaan en -productie, geheugen, aandacht/resourcesallocatie en coördinatie. De kleine steekproefgrootte van de experimenten en het gebrek aan ecologische validiteit betekent echter dat de bevindingen niet altijd overtuigend zijn. Dit doctoraat op artikels presenteert de resultaten van vier kwantitatieve corpusgebaseerde onderzoek over tolken in het Europees Parlement. In totaal zijn 220 tolkprestaties in zeven talencombinaties onderzocht. Er werden tientallen input- en outputvariabelen geanalyseerd, maar het geslacht van de tolk was de belangrijkste controlevariabele voor dit onderzoeksproject. Studies over individuele cognitieve vaardigheden suggereren inderdaad dat vrouwen hoger scoren dan mannen in verschillende verbale taken die betrokken zijn bij de cognitieve processen van het simultaantolken. Deze processen zijn geoperationaliseerd in vier verschillende onderzoeksonderwerpen: de lengte van de Ear-Voice Span, de productie van disfluencies, de weergave van cijfers en het gebruik van extraposities door tolken die in het Nederlands en Duits werken. Voor elk onderwerp is de invloed van sekse geanalyseerd, naast diverse andere input- en outputvariabelen, zoals bron- en doeltalen, sprekersritme en frequentie en duur van stille pauzes. Uit de resultaten blijkt dat geslacht geen significante voorspeller is van de Ear-Voice Span, het gebruik van extraposities en de weergave van cijfers. Mannelijke tolken blijken echter meer disfluencies te produceren, met name gevulde pauzes en verlengingen dan vrouwelijke tolken.

Dit onderzoeksproject heeft het ook mogelijk gemaakt om de bevindingen van eerdere studies te bevestigen of te weerleggen, en om interessante nieuwe inzichten op te leveren in de manier waarop tolken in een echte omgeving werken. De resultaten van de eerste corpusstudie tonen aan dat de lengte van Ear-Voice Span in lijn ligt met eerdere bevindingen, met een mediaanwaarde van drie seconden. Onze gegevens geven ook aan dat de EVS toeneemt wanneer de bronsprekersritme toeneemt. Het Nederlands produceert een langere EVS, wat in lijn is met de bevindingen die erop wijzen dat SOV-bron talen een langere EVS vereisen. De EVS blijkt ook de productie van de meeste disfluencies te verhogen. Zoals verwacht in het tweede corpusonderzoek was de frequentie van fouten en omissies in de weergave van getallen veel lager in onze gegevens dan in experimentele gegevens, aangezien tolken vaak hulp krijgen van hun collega's of de transcript. Onze bevindingen bevestigen dat weglating de meest voorkomende categorie is na volledige weergave. De resultaten vinden echter geen significante invloed van de aard en complexiteit van de cijfers en de bron- en doeltalen op de weergave van cijfers. Het derde corpusonderzoek toont aan dat tolken meer disfluencies produceren dan bronsprekers, wat de neiging heeft om te bevestigen dat tolken cognitief veeleisender is dan algemene taalverdrag. De frequentie van de disfluencies die door de tolken worden

geproduceerd, lijkt toe te nemen wanneer de bronsprekersritme toeneemt, wat in lijn lijkt te zijn met de literatuur, die aangeeft dat de bronsprekersritme de cognitieve belasting verhoogt. De laatste corpusstudie toont aan dat tolken in het Nederlands en Duits hun middenveld meer verkorten dan oorspronkelijke sprekers en extrapositie gebruiken, deels omdat ze de woordvolgorde van de brontaal nabootsen (in dit geval Frans). Deze bevindingen bevestigen dat tolken proberen geheugencapaciteit te besparen. Bovendien verkleinen Duitstalige bronsprekers hun middenveld in vergelijking met hun Nederlandstalige tegenhangers, wat een in de literatuur gedocumenteerde tendens bevestigt.

Dankzij de methodologieën die worden toegepast om het huidige corpus te transcriberen en te annoteren, alsook om de resultaten te analyseren en te kwantificeren, hopen we andere onderzoekers te inspireren om corpusgegevens te gebruiken om cognitieve processen in het tolken te observeren. Meer specifiek moedigen we onderzoekers aan om sekse als voorspellende variabele op te nemen in hun experimentele ontwerpen om onze bevindingen te bevestigen of te weerleggen.

List of Abbreviations

CoSi	Consecutive and simultaneous interpreting
DiK	*Dolmetschen im Krankenhaus* (Interpretation in hospitals)
EPIC	European Parliament Interpreting Corpus
EPICG	European Parliament Interpreting Corpus Ghent
EVS	Ear-Voice Span
SI	Simultaneous interpreting
SOV	Subject-Object-Verb
SVO	Subject-Verb-Object

List of Tables

Corpus Study 1

Corpus Study 2

Corpus Study 3

Corpus Study 4

List of Figures

Chapter 3. Methodological Framework

Corpus Study 1

Corpus Study 4

Table of Contents

PART 1 : INTRODUCTION

Part 1: Introduction

Chapter 1 explores the different theories and models adopted throughout the evolution of interpreting studies, most specifically research into the cognitive processes of simultaneous interpreting. Chapter 2 covers the literature on cognitive sex differences as this dissertation has sex as a main control predictor. Chapter 3 gives an overview of the advantages and disadvantages of past and current methodologies. Finally, the details of the research project are laid out in Chapter 4.

Chapter 1: Theoretical Framework

The expression 'interpreting studies' was first coined in the early 1990s (Salevsky 1993) but the interest in this field dates back to the '60s. At the time, psychologists aspired to understand the processes involved in a relatively new task: simultaneous interpreting, a mode in which the interpreter renders the speech as it is being delivered by a speaker, with a lag time of a few seconds. Indeed, while interpreting is often referred to as the world's second-oldest profession, its current form in most international organisations, i.e. simultaneous interpreting in a soundproof booth, is relatively recent. Simultaneous interpreting was introduced in the late 1920s at the International Labour Organization and during a Congress of Comintern. Its first large-scale use is recorded at the Nurnberg trials in 1945.

In the late 1960s, psychologists and psycholinguists such as Oléron and Nanpon, Goldman-Eisler, Gerver and Barik turned to simultaneous interpreting as a way to better grasp bilingual language processing. At the time, most studies on simultaneous interpreting occurred within the paradigm of experimental psychology, and most experiments were based on behavioural methods, i.e. analysing reactions to different input variables, such as source text type (Barik 1971) or speech rate (Gerver 1969). Goldman-Eisler (1967) realised a pioneer study on silent pauses in simultaneous interpreting and found that pausing patterns are different compared spontaneous speech. In 1975/1976, psychologist David Gerver presented the first cognitive process model of simultaneous interpreting, based on previous experiments. The model presented a system of short-term memory stores for the different stages of speech processing, and also included the role of long-term memory in activating the appropriate linguistic units. Interestingly, Gerver's model presented two buffers, one for the source language and one for the target language, working in parallel. This was very modern for the time, as information processing was still mostly believed to be confined to a single channel. Working memory was also mentioned in Gerver's model but neither its precise nature (structural or functional) nor role was explained. While recent models are more precise when it comes to working memory, none of them has been able to clearly define its role and nature. Moreover, Gerver's model neglects the use of contextual knowledge by the interpreter, as well as the social and communicative context in which interpreting occurs. This approach is sometimes referred to as the 'Conduit' model where the interpreter is seen as a passive conveyor of a message, without playing any role of bicultural-bilingual mediation or communication facilitation. While several other researchers have developed their own model since Gerver, the cognitive process approach focusing on mental processes remains a very popular approach for simultaneous interpreting, given the small role played by social or relational aspects for this mode of interpreting, compared to other types of interpreting such as dialogue interpreting, where the interaction between the interpreter and the other stakeholders is much more present. While modern linguistics had a significant impact on translation studies in the '60s, few researchers had integrated linguistics in the interpreting field, namely because linguistics at the time focused on words or sentences, while the continuous flow of speech interpreters deal with could not easily be divided into clear sentences. However, Chernov (1978) developed the probability prediction model for SI based on discourse-oriented linguistics. This model suggests that simultaneous interpreting is only feasible thanks to the inherent redundancy of discourse, which allows anticipation, i.e. the prediction of source text constituents. It explains why SI is not suitable for dense discourse, such as poetry or dense written documents.

In the 1970s, a new community of actual practitioners of simultaneous interpreting, led by Danica Seleskovitch from the École Supérieure d'Interprètes et de Traducteurs (ESIT) in Paris, started rejecting the frameworks offered by psychology and linguistics, in favour of a socio-cognitive approach. They promoted the analysis of authentic interpretations, instead of data produced in the laboratory, and the use of pragmatic approaches and introspection. They developed their own framework, i.e. the Paris School approach, based on the Interpretive Theory, also called the 'théorie du sens' or interpretive theory in English (Garcia-Landa 1981), which is still largely taught in interpreting schools today. This theory, which was originally meant as a tool to help new teachers of simultaneous interpreting explain their practices to students, suggests that interpreting is not a matter of translating words but of conveying the non-verbal sense of the speaker's message. The interpreting process is divided into three, sometimes overlapping, stages: comprehension (construction of the sense to be transmitted, mostly through the integration of prior knowledge with new information); deverbalization (transformation of the source words into a non-verbal message); and reformulation (the non-verbal sense is transmitted in the target language). Contrary to Gerver's model, this model highlights the importance of external knowledge in interpreting and considers interpreting as an act of communication. However, it neglects the role played by strategies in interpreting, as it assumes that the interpreter only relies on linguistic and world knowledge to understand the original. While this theory is entirely constructed from experience-based intuition and has not been scientifically proven, some support for the theory of non-verbal sense was offered by cognitive psychologists and psycholinguists. Indeed, they have shown that comprehension is based not only on linguistic knowledge but also on prior knowledge and contextual knowledge. Moreover the concept of deverbalization was later supported by evidence from neuropsychological research showing that language and thought are located in different areas of the brain (e.g. Barbizet 1968).

In the late 1980s, a new generation of research-minded interpreters started to question the Paris School approach and trigger a return to empirical research, inspired by Gerver's pioneering work, and largely accepted by the growing international community of researchers in interpreting studies. This new scientific approach led to the adoption of a variety of cognitive models for the interpreting process, mostly based on the information-processing paradigm, which compares the human brain to the central processor of a computer, limited in the amount of data it can process. However, this paradigm assumes that the human brain only has one processor executing one task at a time, while the billions of neurons in the human brain can each be considered as a single processor. Consequently, another cognitive approach called the connectionist paradigm (Rumelhart & McClelland 1986), which sees the human brain as a network of interconnected processors, gained popularity. In 1985, Gile introduced the notion of cognitive load in interpreting as an explanatory variable for information loss. In 1995, he developed the Effort model for simultaneous interpreting, which divides the task into four processes which compete for limited attentional resources (also called processing capacity): listening and analysis, production, memory and coordination. The listening and analysis effort includes the detection and identification of stimuli and the assignment of a meaning to what is heard. The short-term memory effort is presented as a storage mechanism where information is temporarily kept before further processing takes place. The production effort represents the planning and production of the speech in the target language, as well as self-monitoring. Finally, the coordination effort accounts for the management of attentional resources to the three other efforts, given that these resources are believed to be limited. The aim of this model is to lay out the potential consequences of the limited availability of processing capacity, a concept borrowed from cognitive psychology. Indeed, Gile suggested that omissions or errors can only partially be explained by specific difficulties in the source speech (such as accents and technicity) but are mostly due to the fact attentional resources required to perform the task adequately were not available for a particular comprehension, memory or production task at a time when they were needed. The reason behind this lack of availability is that simultaneous interpreters tend to work close to cognitive saturation (referred to as the tightrope hypothesis). Given the limited processing capacity, Gile also mentioned the need for strategies, or coping tactics (such as segmentation, adapting the lag time for example). While this model and the accompanying hypothesis were intended as pedagogical tools and therefore lack rigorous empirical support, they became the most widespread theoretical framework for the cognitive processes in interpreting, probably due to their apparent simplicity. In 1978, Barbara Moser, another research-minded interpreter, presented a new process model of simultaneous interpreting. Her model specifies the nature of working memory, which she calls ‘generated abstract memory’ (GAM), i.e. to be both a structural and functional component. However, she adds that it is identical to short-term memory, as its main role is to store processed chunks of text. GAM is also involved in linguistic transformation, in cooperation with long-term memory, as well as in production. This model, however, is sometimes incoherent (paraphrasing and prediction functions are placed outside GAM) and lacks further explanation given the briefness of the article.

From the late 1990s until today, the psycholinguistic approach of the ’60s is still being applied by several researchers, but with two major improvements. Firstly, the pragmatic dimension put forward by actual practitioners is incorporated into cognitive models, which now take into account input variables, as well as the context and the psychological dimension, contrary to the Conduit model. Indeed, linguistics and more specifically pragmatics have now shown that the use of language is dependent on the context and that the speaker is not the only actor, but that all the present stakeholders have an influence on the communicative process. Researchers start conducting social-interactional research and explore the relation between interpreting and the social context in which it takes place, based on the argument that interpreting is essentially social in nature (Monacelli 2009). Pöchhacker’s (1994) interactant model aims at linking the cognitive and social aspects of interpreting by relating each interpretation to a wider hypertext and to the client-interpreter relationship. Similarly, Setton (1999) applied the Relevance theory to conciliate the cognitive and communication aspects of interpreting. His model suggests that both extra-textual knowledge and pragmatic clues in the discourse are used to decode the message, and takes into account the source speaker’s communicative intent. Monacelli (2000) attempts to restore the ecological validity of interpreting studies by adopting constructivist epistemology as a new approach to interpreting research. This method involves the analysis of naturalistic data and the inclusion of the input of the interpreters who produced the data under study. She uses this approach mostly to better understand the strategies adopted by interpreters. While Monacelli looked at conference interpreting, this approach is more frequently used for interpretation in a social context, because the role of the interpreter as mediator is more visible. Secondly, studies adopt the Expert-Novice paradigm in order to discover the sub-components of the interpreting task. This method consists of testing specific skills between professional interpreters and other language users in order to identify tasks at which interpreters perform better, and which are therefore believed to be involved in SI. Moreover, the expertise of interpreters in specific tasks involved in SI means that interpreters have more mental resources available to deal with the high cognitive load of SI (e.g. Padilla et al. 1995; Christoffels et al. 2006; Bajo et al. 2010). However, the role of certain specific tasks has not been proven yet. For example, while working memory is considered an essential requirement for SI, there is no scientific evidence that superior working memory capacity leads to better performance in interpreting (Timarova et al. 2014). Several researchers have tried to determine whether interpreters have better memory skills than non-interpreters (or novice interpreters) but the results are mitigated. Some studies indicate that interpreters possess better memory skills (Padilla et al. 1995; Darò & Fabbro, 1994; Christoffels et al. 2006) while others have failed to determine so (Nordet & Voegtlin 1998; Köpke & Nespoulous 2006; Timarová 2012) and some obtained mixed results (Stavrakaki et al. 2012). The overall broad conclusion seems to indicate that interpreters do not outperform other individuals, including interpreting students, on simple storage tasks, probably because the average temporal storage period in simultaneous interpreting (the equivalent of the average lag between the source speaker and the interpreter, i.e. about three seconds) is relatively short and rarely goes beyond the short-term storage capacity. Experimental studies with tasks combining storage and processing, on the other hand, provided support for the hypothesis of better working memory in interpreters. It is also important to note that the diverging results might be due to the diversity of methodologies and tests adopted in the above-mentioned studies. Moreover, differences in memory skills between professional and nonprofessional interpreters might not mean that interpreters improved their memory skills through practice, but that good memory skills are a prerequisite for professional interpreters. However preliminary results of a study by Hervais-Adelman et al. (2011) showed an increase in grey matter volume over the course of a 15-month training programme in brain regions known to be involved in aspects of executive function and error monitoring. This can be explained by the fact that training on complex working memory tasks leads to improvement on similar tasks in a specific domain but not on tasks in a different domain (Daneman & Carpenter 1983; Daneman 1991; Liu 2008; Harisson et al. 2013). Consequently, training might improve interpreters' use of working memory in interpreting tasks and help them develop better strategies to enhance their memory skills when interpreting but it does not mean that their working memory skills are better in general. This would call for the actual interpreting skills to be tested in studies on memory, as has already been done by a handful of researchers (Liu 2001; Christoffels et al. 2003; Hodáková 2009; Tzou et al. 2011; Timarová 2012).

The 1990s were also marked by more cooperation between translation and interpreting scholars, notably manifested by a shared interest in corpus-based research. Indeed in 1998, Shlesinger called for more corpus-based interpreting research, inspired by insights offered by translation corpora (Baker 1993). The use of corpora in linguistics has become mainstream and most publications since 2010 are corpus-based. However, the use of corpora for interpreting studies is still that of a niche. This scarcity is partly due to the availability of data. Indeed, the only publicly available data for interpreting studies are those included in the European Parliament Interpreting Corpus (EPIC), a corpus of simultaneous interpreting compiled at the University of Bologna and corpora of various types of interpreting developed at the University of Hamburg: CoSi (consecutive and simultaneous interpreting) and DiK (dialogue interpreting in public service settings). Various initiatives have been taken and are currently underway to create and make available other simultaneous interpreting data drawn from the European Parliament. The Universities of Ghent (EPICG), Posnan, Louvain-la-Neuve and Saarbrücken have recently joined the University of Bologna in an effort to expand the interpreting corpus.

Finally, another area of research is slowly gaining popularity and might change the face of interpreting studies: neuroscience. Cognitive and neuroscientists had already shown an interest for interpreting in the late ’80s, characterised by experimental studies on the lateralization of language, most particularly in case of bilingualism (Gran & Fabbro 1988). Indeed, after conducting neurobiological inquiries, most authors in the ’90s agreed that in monolingual right-handers the left hemisphere is responsible for processing of temporal and analytical information, i.e. the use of language, while the right hemisphere is responsible for processing analogical and spatial information (Broca 1861; Deegener 1978; Damasio 1992). In 1978, Albert and Obler discovered that in bilinguals the right hemisphere is also involved in language, i.e. that bilinguals have a more symmetrical use of both hemispheres. The researchers therefore called for more research on cerebral asymmetries in simultaneous interpreting, since SI was considered an extreme language control task. Daro & Fabbro (1994) were the first to answer this call and to borrow neuropsychological techniques such as verbal-manual interference tasks to study the brains of simultaneous interpreters. Cognitive neuroscience in interpreting today is still dominated by the approach of bilingualism and researchers in interpreting start applying neurological methods to measure the brain activity such as electroencephalography (EEG), which measure the electric activity (Kurz 1996) and functional magnetic resonance imaging (fMRI) (Krik et al. 2005) which measures changes associated with blood flow. These modern approaches have first helped officially confirm the difference between interpreting and other tasks, as well as identify brain areas involved in SI, such as areas associated with cognitive control and speech comprehension (Hervais-Adelman et al. 2014). Finally, they suggest that SI might produce long-term functional and structural changes in the brain (Moser-Mercer 2010) due to the extensive training interpreters receive. Given the complex conditions in which EEG and fMRI methods must be used, psychophysiological methods are also starting to be applied to interpreting studies. They allow for cognitive tasks to be measured through physiological responses such as heart rate, blood pressure or eye movement (Korpal 2016; Seeber 2013). For example, Seeber used pupillometry to test a new model inspired by Gile’s Effort Model, the Cognitive Load model (Seeber 2011). Both models are based on the same principles and the same definition of cognitive load, i.e. a function of capacity and demand, where demand results from the concurrent realisation of two tasks: language comprehension and language production, each consisting of a similar series of sub-tasks. However Seeber offers a more nuanced picture of competing demands and pays more attention to the output load, while Gile’s model is mostly focused on the input load. Contrary to Gile’s model, Seeber’s model is also meant to be predictive and testable.

While advances in interpreting studies are significant, some areas remain to be explored. For example, while several strategies have been identified, the consciousness of their use is still a matter of debate. Moreover, the numerous useful models for cognitive processes of SI are still tentative and there is still not enough scientific proof to determine which model or which approach is more suited to SI. This is mostly due to the complexity of the interpreting task, the inter-subject variability, as well as the difficulty to guarantee ecological validity (Setton encyclopedia models). Some other topics still under study include the executive function of working memory, the management of the cognitive load and the precise influence of input and output variables on interpreter performance. The use of more advanced methods borrowed from neuroscience might help find answers.

Chapter 2. Sex as a predictor of interpreting

1. Sex differences in cognitive skills

Three decades ago, Gran & Fabbro (1988) concluded their study on interpreting by stating that women are neurobiologically better suited for interpreting tasks. While today this statement would probably be highly disputed, researchers at the time considered that women were at an advantage in language tasks due to a more symmetrical cerebral representation for language functions (Gran & Fabbro 1988). Numerous studies have reported significant differences between the sexes relating to brain structures and the ability to perform, specific cognitive tasks. In recent years, however, many researchers have grown sceptical about claims of cognitive sex differences. Consequently, several meta-analyses have been conducted as a way to obtain a broader picture. This Section presents an overview of the current state of research on cognitive sex differences.

Neuroscientists have long been interested in exploring sex differences in the human brain to explain the observed differences regarding prevalence of psychiatric disorders, such as Alzheimer for females (Mazure & Swendsen 2016) and schizophrenia for males (Aleman et al. 2003), or certain some psychological traits, such as physical aggression for males (Archer 2004) and agreeableness for females (Costa et al. 2001). Unfortunately, it has been noted that the potential influences of sex are under-explored in neuroscientific research (Cahill 2017). Moreover, studies on sex differences in human brains tend to show contradictory results, which may be due to small sample sizes (Nord et al. 2017) and/or variability in age range of the sample in individual studies (Ritchie et al. 2018). While several meta-analyses have been conducted recently in an attempt to find conclusive results, sample sizes remain insufficient for statistical significance of differences to be tested (Ruigrok et al. 2014) or the focus is on young participants only (Gur & Gur 2016; Gennatas et al. 2017; Wierenga et al. 2017). Moreover, when sample size is sufficient, most studies tend not to adjust these differences for overall brain size, which usually renders the observed differences nonsignificant, and there always seems to have a strong overlap between males and females (Ritchie et al. 2018). Fortunately, meta-analyses on more specific areas of the brain have enabled the following two conclusions to be drawn: No sex differences are found in amygdala volume (Marwha et al. 2016), nor with regard to the hippocampus (Tan et al. 2016). Gran & Fabbro (1988) argue that women are better suited to perform the task of interpreting, citing differences in lateralization. Indeed, while many studies have investigated sex differences in language lateralization, the results are not entirely conclusive. However, recent studies, including a meta-analysis (Sommer et al. 2008) have concluded that there are no sex differences regarding asymmetries and lateralization (Chiarello et al. 2009).

If differences in brain volumes or surface areas do exist, they do not automatically translate into differences in behaviours or cognitive abilities, given the other factors that might come in play. Indeed, researchers found weak correlations between brain variables and cognitive tests, and these associations did not differ by sex (Schnack et al. 2014; Pietschnig et al. 2015). Some researchers also suggest that the structural differences of the brain can sometimes be the result of compensatory mechanisms for differences in sex-specific hormones, and might thus actually attenuate behavioural sex differences that would otherwise have been present (De Vries 2004; McCarthy & Arnold 2011).

Several meta-analyses have been conducted to assess the existence of sex differences in general cognitive tasks. Results indicate that males and females are much more similar than they are different and that sex differences are often exaggerated (Hyde & Linn 1988; Hyde 2005; Miller & Halpern 2014). When it comes to sex differences in verbal abilities more specifically, Hyde & Linn's meta-analysis (1988) also shows that the magnitude of sex differences is currently so small that it can effectively be considered negligible. While meta-analyses are essential for determining whether sex differences exist as a whole or with respect to general cognitive abilities, Hyde & Linn suggest that we study sex differences in abilities more precisely and 'move away from the old model of intellect that specified only three rather general cognitive abilities – verbal ability, mathematical ability, and spatial ability' (p 33). While Chapter 2 aims to illustrate the complexity of cognitive processes modelling for simultaneous interpretation, researchers agree that it comprises the following tasks: speech comprehension, speech production and memory. Accordingly, studies on the influence of sex in tasks linked to these three main cognitive processes will be explored. It has to be mentioned that while studies indicating sex differences on specific tasks appear to be quite numerous, most of them unfortunately do not mention effect sizes (nor provide sufficient information for these to be calculated by the reader). Moreover, when effect sizes are mentioned, they are generally small, ranging from 0.3 to 0.6.

When it comes to speech production, sex differences have been found in verbal fluency. Verbal fluency is the ability to verbally generate specific information within restricted search and time parameters, and is believed to play a key role in SI (Stavrakaki et al. 2012). Herlitz et al. (1997) conducted a study on a thousand individuals, aged 35–80 who were asked to generate as many words as possible beginning with the letter A in one minute, as well as to five-letter words beginning with letter M. The results show that women outperform men on both tests. Mailtand et al. (2004) carried out a similar study on a larger sample (1796 participants aged 35 to 85 years old), where participants were asked to generate as many words as possible for one minute for three different criteria. Loonstra et al. (2001) also found higher scores for women in verbal fluency tasks after collecting results from the Controlled Oral Word Association Test (COWAT), which is a measure of a person's ability to make verbal associations to specified letters. In a corpus-based study of telephone conversations of more than 600 speakers, Shriberg (1996) finds that men produced more filled pauses than women. Similarly, Bortfeld et al. (2001) show that men present higher rates of filled pauses and repeats in a corpus of task-oriented conversations. However, Engelhart et al. (2010) did not find significant differences between the sexes when studying participants with attention deficit/hyperactivity disorder. While filled pauses can facilitate comprehension in some cases (Corley & Hartsuiker 2011), they are often considered as a consequence of the cognitive load.

When it comes to speech comprehension, sex seems to influence the phonological input processes activated when the human brain perceives a spoken word and tries to identify it. Indeed, in a study on seventy-one subjects, Aerts (2013) found that women display a larger sensitivity to the phonemic contrasts during auditory phoneme discrimination (i.e. the discrimination between different speech sounds in order to identify the spoken word) and showed more differentiation in real word-pseudoword dissociation.

For the speech comprehension tasks, studies have found that girls between 6 and 16 score higher for the subtest coding on the Wechsler Intelligence Scale for Children (WISC), an individually administered intelligence test for children (Lynn et al. 2005; Van der Sluis 2006). The Coding subtest is a measure of processing speed (the speed at which a person can understand and react to the information they receive) and consists of presenting the individual with a key in which the numbers 1 to 9 are each paired with a different symbol; his/her task is then to use this key to put in the appropriate symbols for a list of numbers between 1 and 9. The test is also affected by other cognitive abilities such as learning, short-term memory and concentration. Camarata & Woodcock (2006) analysed the results of three large-scale studies and found that females scored significantly higher on estimates of *Gs* (processing speed) in all three samples. In another review of recent large-scale studies, Roivainen (2011) conclude that women have an advantage with regard to processing speed involving letters, digits and rapid naming tasks. A female advantage in both prelexical and lexical processing was found (Majeres 1999), as well as for perceptual speed (the ability to compare or recognize items) (Born et al. 1987; Hedges & Nowell 1995).

Yet, the specific nature and functions of the human memory are still unclear. Traditionally, long-term memory has been believed to be divided into two categories: declarative memory (memories that can be consciously recalled as facts and verbal knowledge) and procedural memory (unconscious memory as for example automated tasks and skills). Declarative memory is also divided into two subcategories: episodic memory (personal experience, autobiographical memory) and semantic memory (factual information, knowledge of the world). Memory supposedly works in three stages: encoding (receiving, processing and combining received information), storage (creation of a record of the encoded information) and recall (retrieval of information). While the interpreting-specific role of memory is also relatively ambiguous, it is believed to comprise processes involved in acquiring information, retaining said information for a period of time and its subsequent retrieval when needed (Bajo & Padilla 2015). These processes are predominantly linked to short-term memory. Memory is also believed to be involved in various cognitive processes, via its most complex component: working memory (Baddeley & Hitch 1974). While short-term memory serves to retain information over short periods of time, working memory is responsible for maintenance of information (actively refreshing the information so that it is not forgotten) during processing (manipulation of information – e.g. its translation into another language) (Atkinson & Shiffrin 1971; Baddeley & Hitch 1974). As mentioned by Baddeley (1996), long-term memory assists working memory by providing background knowledge (through declarative memory) as a means to facilitate the retrieval of information and the rapid comprehension of the input. Studies by Ericsson and colleagues (Ericsson & Kintsch 1995; Hu & Ericsson 2012) also suggest that expert performance relies upon the availability of well-organized declarative knowledge structures linked to working memory, referred to as "long-term working memory" (LTWM). These structures are believed to develop over years of deliberate practice and evidence supporting the use of LTWM processes by interpreters was provided by Padilla et al. (2005) and Christoffels et al. (2003).

Herlitz et al. (1997) conducted a study on a thousand individuals, aged 35–80 and found that women consistently performed at a higher level than men on the episodic memory tasks. Kramer et al. (1997) studied 811 subjects aged 5 to 16 and concluded that girls perform better on all of the immediate and delayed recall tasks (retrieval from memory without a cue) and on the delayed recognition tasks (retrieval from memory with a cue). Girls displayed more effective long-term memory mechanisms. Boys made more intrusion errors and displayed greater vulnerability to interference between the 2 tasks. When it comes to differences in adults, Herlitz et al. (1999) studied 200 participants aged 20 to 40 and found that women performed better on most verbal episodic memory tasks and on some episodic memory tasks with a visuospatial component. The addition of this visuospatial component was to prove that women's higher performance on episodic memory tasks cannot fully be explained by their superior performance on verbal production tasks. He found larger difference for free recall than for recognition, which he argues is probably because free recall needs verbal production, which women are better at. Maitland et al. (2004) reproduce the study but with a larger sample (1796 participants aged 35 to 85 years old) and conclude that there is a female superiority in semantic memory and a female superiority in declarative memory. Several studies have also shown that women perform better than men on immediate and delayed free recall and on recognition tasks with verbal and visual components (Ruff et al. 1989; Trahan & Quintana 1990; Wiederholt et al. 1993; Kimura & Seal 2003). However, Harness et al. (2008) report higher scores for males in a study on recall combined with a distraction task carried out on students. Other studies report that females outperform males in the Rey

Auditory Verbal Learning Test (free recall of two lists of nouns, which aims at evaluating short-term auditory-verbal memory, retroactive, and proactive interference, retention of information among others) and the Verbal Paired Associates test (immediate and delayed recall of word pairs, aimed at evaluating explicit episodic memory performance) (Bolla-Wilson & Bleecker 1986; Bleecker et al. 1988; Geffen et al. 1990; Gale et al. 2007).

2. Sex in interpreting studies

The dimension of sex is barely explored in interpreting studies. Moreover, when sex is mentioned, it is more often in reference to gender. Before going further, it is therefore important to clarify the difference between sex and gender differences. Although the words gender and sex are often used interchangeably and the expression "gender differences" is much more frequent (sex is typically applied to sexually dimorphic traits), they mean different things. Sex refers to a person's biological sex, i.e. the anatomy of his or her reproductive system and secondary sex characteristics). While sex is mostly considered to be binary (male or female), several individuals cannot be placed in one of the two categories, as they have biological sex characteristics that complicate sex assignment, and might be intersex, i.e. born with variations in chromosomes, gonads, sex hormones, or genitals that do not fit the typical definitions for male or female bodies, which concerns about 1.7% of the world population. The issue of gender is a bit more complex as it can refer to two different concepts: either a person's gender refers to the social role expected by society, based on the sex of the person, or to their personal identification of one's own gender based on an internal awareness. While traditionally, the female sex is assigned to the gender female, a person's sex and gender do not always align, and the person may be transgender, i.e. have a gender identity or gender expression that differs from the sex that was assigned to them at birth. Moreover, some people do not identify to one of the two traditional genders, but consider themselves non-binary, bigender or genderfluid among others. This distinction between sex and gender is not globally accepted though and several dictionaries don't make a distinction between gender and sex. When it comes to research on differences between males and females, some researchers disagree with the distinction made between the terms sex difference and gender difference and argue that an individual's behaviour cannot be partitioned into separate biological, developmental and cultural factors because it is not always clear the degree to which the differences are due to which factor (Halpern 2000; Lippa 2005).

In the present dissertation, reference will be made to the interpreter's biological sex and to sex differences, as the subjects are anonymous and no assumptions can be made on their gender. This does not mean that, if differences are found for the topics in question, these are not also due to social factors related to the gender of the interpreter under study. Moreover, for the reasons mention before, the term "sex differences" does not appear in the title of this research project. The other reason being that in our corpus studies, sex is a predictor among other predictors and not the only predictor under study. Finally, the term "differences" gives the impression that authors are digging to find differences.

As mentioned above, studies on sex differences in interpreting are very rare. In small preliminary corpus-based studies, Defrancq (2013) and Baes (2012) discover longer

EVS for female interpreters at the European Parliament. In an experimental study on 11 professional interpreters, Cecot (2001) finds that women produce an average of 10.7% more filled pauses than men and men produce an average of 14.9% more silent pauses than women. Other studies tend to take the concept of gender more into account and look at difference regarding the individual's self-concept and role within society. In corpus-based-studies, Magnifico & Defrancq (2016, 2017) have discovered that female interpreters at the European Parliament use more hedges and downtone fewer unmitigated face-threatening acts than male interpreters (with large effect sizes, respectively Φ=0.4 and 0.49). Analysing 704 answers from professional interpreters to a survey on the perception of quality, Pöchhacker & Zwischenberger (2010) show that female interpreters rate others' performances more generously and value a lively intonation in interpreting more than male interpreters. However, Angelelli (2004) found no gender difference in the analysis of interpreters' perceptions of roles and behaviours in practice. Analysing recorded courtroom proceedings, Mason (2008) looks at the sex and gender dimension of discourse markers in interpreting, but with a stronger focus on gender. She finds that female consecutive interpreters omit linguistic features that signal deference more than men, while male interpreters omit more politeness markers. As can be quite easily understood, these findings focus on specific populations such as interpreters at the European institutions or interpreters in courts and can therefore not be generalised to all interpreters.

3. Potential implications

As previous sections show, sex is not a popular topic in interpreting studies. Interestingly, the review of literature of sex differences in other fields with a much longer tradition of research seems to indicate that they also suffer from a lack of large scale and replicable studies including sex as a predictor. This lack of scientific evidence is worrisome as it can push researchers to trust findings based on insufficient data and draw inappropriate conclusions. For example, Gran & Fabbro were convinced at the time that sex differences in lateralization did exist, while recent research shows that this tends to be untrue. With a lack of conclusive findings, it is very tempting to make false statements. Moreover, the general confusion around the concepts of sex and gender mentioned in the previous section show that this topic needs further research and attention. The recent inclusion of neuroscientific technologies in interpreting studies is certainly good news as it will highly contribute to literature on the topic in a task that involves many of the areas for which theories are still very tentative (e.g. working memory and resource management). But as mentioned above, differences or similarities in brain structures or volumes are not sufficient and the actual behavioural differences must also be studied. Moreover, the issue of small sample sizes has been mentioned in studies on sex differences, and is certainly even more worrying for studies in interpreting, given the small population. This issue is most probably not going to be resolved in a very near future and a partial complement can be offered by corpora. Indeed while corpora have several limitations, they allow for larger groups of interpreters to be analysed. While this research project will not solve the issue of a lack of scientific evidence on its own, the authors hope that it will inspire more researchers in interpreting studies to include sex as a predictor. The aim is not to fish for differences where they don’t seem to appear, but simply to gather more data in order to determine whether they exist or not and in which aspects of interpreting.

Indeed the existence or absence of sex differences can have several implications. As the previous chapters show, some gender differences have been found in interpreting. While these studies are also quite limited in size and representation, they could be complemented by studies on sex differences. Indeed as mentioned in the introduction, behavioural differences are often caused by a series of factors, and having both sides of the medal, nature and nurture, can help have a larger picture. Consequently, findings from studies on sex differences can give complement studies on differences found in the role of self-conception of interpreting. On the other hand, when no cognitive sex differences are found in places where gender differences do appear, this could suggest that the observed differences are mostly societal, cultural or educational and are not based on cognitive differences.

While exceptions do exist in particular countries or contexts, statistics show that there is an imbalance between the sexes when it comes to the profession and to interpreter training around the world, female students outnumbering male students to a considerable extent (Lim 2005; Ryan 2015; Hickey 2018). This situation is a popular topic among interpreters, proof is the latest conference organized by ISIT (Discuss interpreting) where a presentation by Sarah Hickey (2018) on reasons for a female predominance in interpreting stirred a lot of discussion in the room and online. While participants had their personal theories, the reasons for the predominance of women in the profession remain unclear. Hickey and Ryan provide some food for thought in their master's thesis. No scientific study has been conducted on this topic. However this phenomenon potentially has worrying origins and consequences. Miller & Halpern (2014) suggest that individuals' relative cognitive strengths are important to their career and educational choices. Studies on cognitive sex differences could therefore help explain why fewer men opt for a career in interpreting, while potentially eliminating stereotypes. Studies on sex differences in verbal tasks showing a female advantage might have an impact on men's decisions, as men might not choose interpreting based on their potential impression that having a female predisposition is a prerequisite for a career in. This perception could also have an impact on their performance, given that people's performances appear weaker when they are told that the other sex performs better at the task (Spencer et al. 1999).

Finally, if sex differences do exist, they could be taken into account when designing interpreter training. Indeed studying sex differences is considered an important step to better understand what drives differential performance in professionals and students and help design effective training. Halpern et al. (2007) suggest that we can use this knowledge to teach female and male students ways to solve problems that correspond to their most efficient cognitive process to allow more flexibility in their problem solving and positively impact performance overall. While it is unlikely for sex-specific courses to be organized, sex differences would suggest that trainers need to take individual skills into account instead of offering one-size-fits-all solutions.

Chapter 3. Methodological Framework

1. Methodology

As mentioned in Chapter 2, several approaches have been adopted in the course of the development of interpreting studies. Each approach comes with specific methodologies. This Chapter is an overview of past and current methodologies through several of the most iconic studies on SI.

As mentioned in Chapter 2, experimental psychology has been a popular framework in interpreting studies since the ’60s. At the time, researchers were mostly interested in analysing reactions to different input variables. For example, the psychologist David Gerver (1969) dedicated his doctoral dissertation to the influence of the input speech rate on interpreters’ performance. The subjects of his experiment were 10 professional interpreters working from French into English. They were divided into two equal groups: interpreters and shadowers. Shadowing is a method often used in comparison with SI, even today. Indeed, shadowing resembles the interpreting process, as it consists of listening and speaking at the same time. The difference, of course, is that shadowing is monolingual and interpreting is bilingual. Both groups deal with one source speech with an original speech rate of 120 words per minute, which has been artificially sped up and slowed down in order to obtain 5 different delivery rates (from 95 w/m to 164 w/m). Gerver’s hypothesis was that the two groups would react differently to the different delivery rates, given the difference in their tasks. He also suggested that interpreters would take advantage of pauses to catch up with the speaker. He reached the following conclusions: the two groups have similar behaviours when the rate is under 110 words/min. Above that, differences between the two groups are observed: shadowers increase their speech rate as the source rate increases, while interpreters seem to reach a ceiling of 110 words/min. Secondly, interpreters increase the lag time between them and the source speaker when the source speech rate increases, while shadowers keep the same lag time. There are some similarities between the two groups as well: both have fewer silent pauses when the rate increases. While results are interesting, this experiment lacks ecological validity. Indeed, a speech that has been artificially sped up or slowed down is not going to have the typical features of naturally slow or fast speeches. Despite this shortcoming, this experiment is often cited today when referring to the influence of source speech rate on interpreting.

Later, psychologists turned to more advanced methods borrowed from neuropsychology. Gran and Fabbro (1988) for example used the dichotic listening tasks and the tapping task with interference in an attempt to determine whether interpreters, i.e. subjects with a strongly developed second language, use both brain hemispheres when dealing with languages. Indeed, each hemisphere is believed to be linked to different functions: in monolingual right-handed people, the left hemisphere is responsible for the processing of temporal and analytical information, i.e. the use of language, while the right hemisphere is responsible for processing analogical and spatial information (Broca 1861; Deegener 1978; Damasio 1992). But previous studies on bilinguals have shown that they use both hemispheres. The dichotic listening task is used for the first time by Kimura (1961) and consist of sending two different acoustic signals simultaneously in both ears and determining whether subjects respond differently to each signal. Since researchers believe that the right ear is linked to the left hemisphere and the left ear to the right hemisphere, this method is used to determine which hemisphere is more mobilised for a given task. Given that interpreters working in simultaneous interpreting usually cover only one ear with the earphone from which the source speech comes and listen to their rendition with the other ear, this task is highly related to SI (Lawson 1967; Lambert 1989). The tapping task with interference developed by Kinsbourne & Cook (1971) is also aimed at determining which hemisphere is mobilised, but consists of subjects pressing a button as fast as possible, either with their right index (believed to be controlled by the left hemisphere) or left index (believed to be controlled by the right hemisphere), while performing a concurrent task. For Gran & Fabbro's experiment, three groups of twelve students with Italian as mother language were formed. The first group had a good knowledge of English, the second group consisted of student interpreters with excellent knowledge of English and the third group had a rudimentary knowledge of English. For the dichotic listening tasks, students were asked to remember lists of numbers in English and Italian, where different numbers are sent simultaneously to each ear. There was a statistically significant difference between both ears for all groups, i.e. more numbers were recalled from the right ear (linked to left hemisphere, responsible for language), with the exception of English for the group in student interpreters, where no difference between both ears is found. For the tapping experiment with interference, a group of 14 right-handed students interpreters had to recite the days of the week in Italian and in English while either tapping with the right or left index. The same experiment was performed but students had to interpret simultaneously while tapping. It appeared that none of the language was lateralized, contrary to experiments with monolinguals, who usually score higher with the right hand. Gran & Fabbro conclude that the intensive training in English that the group of interpreters has been through has enhanced the use of the right hemisphere for language. However, it has to be mentioned that the fact that interpreters are more able to deal with this task (i.e. tap with both hands and recall numbers from both ears) might not be due to a more symmetrical used of hemispheres, but merely to the fact that they are listening to two different input channels at the same time and have therefore more ease with the concurrent tasks. Moreover, while this type of experiment triggers significative results, Gran & Fabbro add that experimental inquiries into specific aspects of cerebral language functions cannot provide answers to highly sophisticated questions on simultaneous interpreting as yet.

In the late ‘90s, actual practitioners of simultaneous interpreting started applying modern techniques of neuroscience to study their practice. This led to a series of case studies with promising results for the future. Kurz (1996) carried out a case study with electroencephalography to analyse the potential quantitative and qualitative differences in brain activity between interpreting and other tasks. The experiment was carried out on four professional interpreters interpreting from English into German and vice versa. Subjects had to mentally interpret, shadow, listen to music and perform mental arithmetic. These tasks are only performed mentally because EEG is too vulnerable to noise generated by articulatory movements. The results show that there is a difference between interpreting and other tasks, and a more significant use of the right hemisphere. While the conclusions are very tentative given the sample size and the lack of previous research, they show that using EEG during mental interpreting is a viable option, that is worth exploring, as the effect of a stimulus can be directly observed. However the responses cannot be precisely located, as they can be the result of different activities. Krick et al. (2005) turned to fMRI to study brain activity in code-switching and inhibition. The subjects were divided in several groups: one group with limited language skills, one group with good language skills and a third group of interpreters. Subjects had to read a story in which the language changed after every third sentence. Results show that when interpreters switch from one language to the other, a specific area of the brain (Brodman area 46) is activated (while it is not in other subjects). Krick et al. also noticed an increased amount of grey matter in the corresponding area. These findings suggest that training and expertise in complex skills, such as SI, might result in changes in the brain. As mentioned in Chapter 1, these findings were later corroborated by Moser-Mercer (2010) and Hervais-Adelman et al. (2011). Following up on the study by Krick (2005), Ahrens et al. (2010) carried out a new experiment using the fMRI technique in order to investigate the potential differences between 6 German-speaking student interpreters and previous subjects. Students were first asked to talk freely in German then interpret from Spanish into German. Results show statistically significant differences between the two activities. The experiment also identified areas activated by SI: the superior temporal sulcus, which plays a role in the processing of lexical and semantic information, and more surprisingly the activation of hand movements. More research is indeed needed in order to offer more detailed explanations for these results. Contrary to EEG, fMRI allows for precise location of brain areas stimulated by certain tasks. However, the latency required for the stimulation makes it impossible to relate specific events in the input or output to a particular activation pattern, and this method can therefore only be applied to single words or sentences.

Researchers also turned to the output of interpreters to study interpreting. In this case, two main methods are adopted: looking for errors and omissions in the interpretation (Barik 1975; Mazza 2001) or rating the interpretation on a series of criteria and user expectations (Kurz 1993). Errors are mostly used in order to try to understand the processes that have led to these mistakes. Barik (1975) is one of the first researchers to analyse errors in interpreting in a systematic way. The aim of his study was to measure the effect of several speech types (spontaneous, semi-prepared, prepared for oral presentation and prepared for written preparation) by comparing the interpretations into several languages produced by two professional interpreters, two student interpreters and two bilinguals. Professional interpreters had the fewest omissions, followed by students and bilinguals. When interpreting formal lectures, errors are higher for professionals because they omit less. Another classic example of this method is applied by Gile (1999) when testing his Tightrope hypothesis. While the Effort models are not meant to be tested, as they are not predictive theories but conceptual frameworks, the assumptions associated with the model can be tested. One of the assumptions is the fact that errors are an integral part of interpreting and that the many errors that interpreters make cannot simply be explained by traditional source speech factors but are mostly due to the fact that interpreters work beyond their cognitive capacity. In order to prove this theory, Gile carried out an experiment with 10 seasoned interpreters working from English into French and asked them to interpret the same speech twice. His hypothesis is that the second interpretation should contain fewer errors than the first, or at least not include new errors. The results show that errors and omissions occur at different places the second time, and that the majority of interpreters make new errors in the second interpretation. Gile therefore concludes that his theory was right: errors are not linked to specific difficulties in the source speech, but to a lack of available cognitive resources at a specific time. Analysis of the data from one of Gile's experiments indicates that the cognitive management of omissions is indeed highly variable. Omissions that are low-risk for the aims of the speech occur in a constant background mode, almost without source-text stimuli, such that they are found in repeat performances with similar frequency but in different places. However, Pym (2008) re-analysed the data from Gile's study and suggested that when a distinction is made in the types of errors and omissions made by the interpreters, the results do not automatically confirm the Tightrope hypothesis. Indeed, Pym (2008) divides omissions into two categories: those incurring low levels of risk for the interpreter, and those incurring high levels of risk, i.e. hampering the communication. He noticed that in the second interpretation, high-level risk omissions tend to be repaired by the interpreter, and the new omissions are low-risk. This indicates that the interpreters did actually mostly improve their second rendition. While Pym's analysis yields different results and calls for more contextual factors to be included in studies on errors and omissions, the method adopted by Gile, i.e. repeat performances, is very useful. Indeed, they allow for a distinction to be made between what the interpreters do when they have added capacity (i.e. the second version).

A few decades after Gerver's experiment, Shlesinger (2003) analysed the impact of speech rate on interpreters' performance, more specifically on their short-term memory.

Her aim was to test the interpreters' capacity to retain long left-branching noun phrases (i.e. a noun preceded by a long string of adjectives) while interpreting into a head-initial language (i.e. one which requires that the noun be produced before its modifiers) when dealing with different input rates. She formulated two seemingly contradicting hypotheses, based on contradicting findings from previous experiments on recall. On the one hand, she expected interpreters to render more items if the input rate was faster, as less time will elapse between the sounding of the source input and its reconstruction in the target output. On the other hand, the rendition of items in the list was expected to be poorer at the higher rate because it gives the interpreter less time to retrieve the equivalent items from long-term memory. Sixteen professional interpreters working into Hebrew from English and into English from Hebrew, took part in the experiment. Subjects were asked to interpret the same source speeches at two different speech rates (120 w/m and 140 w/m) with a three-week interval. Results show that in one third of cases, interpreters omitted the whole list, and in another third of cases, interpreters omitted three items out of 4-four. There is no statistically significant differences between the two input rates. Shlesinger was probably not expecting such a high frequency of omissions, but concluded that it was probably due to the inherent difficulty of the task and the prioritisation by the interpreter, i.e. the choice to retain and discard some information. She mentions that being able to take that decision is crucial to adequate performance of the task. Indeed, as mentioned before, trying to draw conclusions based on errors and omissions only can be misleading as it neglects the role played by strategies. In order words; the interpreters might have had the sufficient memory capacity to render the items but decided that they were not essential and discard them in order to keep the capacity for the yet-to-come segment. Moreover, given that the lists were artificially added to the test, they do not seem to add highly relevant information to the message. As mentioned by Pym (2008), most of the omissions could probably be considered low risk. Actually, in interpreting training, students often learn that lists are not to be prioritised and that they can remove several items from a list without hampering the communication.

In 1978, Marianne Lederer analyses the output of an interpreter working from English into French based on a 36-second corpus example. She concludes that pauses and a long time-lag can only be explained by the need for the interpreter to understand the message, especially if silent pauses occur when the source speech is not dense. She also notices that previous knowledge from former conferences is spontaneously incorporated. Finally, she notices the use of anticipation twice. Today, anticipation is one of the most popular strategies in interpreting studies, probably due to the simplicity to detect it. Inspired by Lereder's study, Van Besien (1999) studied anticipation in a corpus collected by Lederer (1981) during an international meeting where two interpreters work from German into French. On top of the actual recordings, Van Besien asked the two interpreters from the recordings to interpret the parts of the meetings that they did not interpret during their assignment (since interpreters work in thirty-minute turns). She counts the occurrences of anticipation in the recordings from the conference and after the conference and finds 78 anticipation for 110 minutes of recordings, i.e. one anticipation every 85 seconds. Moreover she finds no significant difference between the two interpreters in the frequency of anticipation. In contrast, Defrancq (2013) finds only one anticipation in a corpus of 150 minutes of recordings for interpretation from French into Dutch. The high frequency of anticipation might be due to two factors: interpreters in her study work from a Subject-Object-Verb language (German) into a Subject-Verb Object language (French), which means that interpreters risk overloading their memory if they wait for the end verbal group. Anticipation is therefore a popular strategy. Indeed when looking at the type of anticipation, it occurs that more than 75% of cases of anticipation concern the final verb. It might be argued that this ‘strategy’ is not an interpreting-specific strategy but a strategy generally applied by all users of Subect-Object-Verb languages. Secondly, half of the recordings concern interpretations that have been performed by an interpreter who has heard its boothmate’s interpretations, as well as the source speech. Interpreters might therefore have remembered the speech, which facilitates anticipation. More than 60% of cases of anticipation are proper rendition of the source items, while the rest are approximations.

2. Experimental and corpus-based approaches

As this review of methodologies in interpreting studies shows, experiments are the most frequent type of research. Experimental designs allow for specific variables to be isolated in order to identify cause-and-effect relationships and inferences about mental processes. Moreover these pioneer studies have delivered great methodological input and hypotheses for future research. More specifically, the increased used of modern techniques, such as EEG and fMRI is very promising.

Unfortunately, all studies mentioned above have small very sample sizes. Researchers often mention the difficulty of finding volunteer professional interpreters to carry out experiments. Given that researchers often work at universities, they often recruit their students for their experiments. While the output of students is perfectly valid, several differences have been found in performance between student and expert interpreters. These indicate that findings about students cannot always be generalised to professional interpreters. Moreover recruiting interpreters or students on-site often means that researchers can only study the languages most frequently used in their region. Given the role played by language pairs in interpreting, this also limits the generalisation of findings. This lack of volunteers and diversity is one of the main reasons researchers turn to corpus data.

Moreover, in order to yield reliable and relevant results, the different parameters of experiments must be carefully designed (source texts features for example) and the conditions need to be rigorously controlled. However, these conditions often mean that the experimental setup is completely removed from real-life interpreting. Indeed, source texts and speech rates are artificially manipulated (Gerver 1969) and often pre-recorded. Moreover, experiments do not allow for the communicative purpose of interpreting to be achieved, as the subject often interprets for the researchers and not for someone who actually needs the interpretation. By doing so, the interpreter might adopt a more scholar approach without fully playing his role as a cultural mediator. Finally, subjects often work alone and not with a boothmate, which is very unusual for simultaneous interpreting where interpreters always work in pairs. Moreover, interpreters often help each other when interpreting and have therefore developed specific reflexes that they cannot reproduce when working alone. This is particularly true for the rendition of numbers, where interpreters almost always receive help from their boothmate. Given these shortcomings, Shlesinger (1998) called for more corpus – based research to be carried out in order to restore the ecological validity and complement findings from experiments.

As mentioned before, looking for errors and omissions is a popular method in experiments. While this approach can yield interesting results, it tends to be too strict and omit strategies linked to errors. Moreover, the methodology behind this approach is not always entirely solid. For example, Gile (1999) instead of looking at the link between cognitive processes and errors, tries to prove the absence of a link between source textual properties and errors in order to conclude that if the text is not the reason for errors, then errors are due to the difficulty of the task. Unfortunately, this reasoning is valid only if errors were triggered exclusively by textual difficulties and cognitive constraints. Pym (2008) mentions this tendency of researchers in simultaneous interpreting to only look at these two factors and suggests that other factors play a significant role: aims of the discourse and the strategies of the speakers, for example. Shlesinger (2003) actually mentions the lack of ecological validity of her material and quotes Setton (1998) mentioning the need for the information processing community to test theories on corpus data.

Corpora offer the advantage of reflecting the interpreting activity as a socially situated activity in a way experimental data cannot (Shlesinger 1998). Moreover, data can be collected from a variety of different contexts and individual language users in order to offer a more comprehensive view on interpreting. Unfortunately corpora in interpreting studies are still at the status of nano-corpora given their small sizes. Moreover, they are mostly based on recordings of plenary sittings at the European Parliament, which means that they cannot be generalised to the rest of interpreters.

Researchers in favour of corpus-based research in interpreting expect that they will grow sufficiently and bring to our field the same amount of insights they provided to translation studies. Indeed corpus-based research into specific textual properties of translation has contributed enormously to our understanding of translation. The same can be expected of interpreting corpora. It is also important to note that corpus-based approaches can increase cooperation between translation and interpreting researchers since corpus-oriented methodologies are well established independently from the nature of the corpus data under analysis, results can be compared across disciplines. Also the current availability of corpora of non-mediated spoken data means that findings from interpreting corpora can easily be compared to non-mediated language.

Chapter 4: The research project

1. Introduction and research questions

Simultaneous interpreting (SI) is a complex and cognitively demanding task (Gerver 1969; Barik 1975; Gile 1995; Seeber 2011, 2013), involving numerous cognitive processes (speech comprehension and production, memory, attention/resource allocation and coordination) and is influenced by a plethora of factors. This diversity and complexity make interpreting a fascinating research topic. While research in SI dates back to the sixties, some aspects of interpreting studies remain hardly explored. This research project aims to explore two of them: sex as a predictor and advanced corpus methods for exploratory studies.

Given the complexity of the task, training and continued practice play a crucial role in interpreting. In order to help students better understand the interactions and implications of the cognitive processes involved in simultaneous interpreting, Gile (1995, 2018) divides them into four efforts: the listening and analysis effort, the short-term memory effort, the production effort and the coordination effort. The four efforts compete for limited processing capacity and interpreters need to find a balance between these simultaneous cognitive tasks. Strategies (e.g. anticipation and segmentation among others) therefore play a key role in interpreting, as they allow interpreters to manage the limited cognitive resources allocated to each effort, especially memory capacity, and to deal with additional difficulties (e.g. numbers). Gile (1995) also suggests that omissions or errors are caused not only by specific difficulties in the source speech but because attentional resources required to perform the task adequately were not available for a particular effort at a time when they were needed. The reason behind this lack of availability is that simultaneous interpreters tend to work close to cognitive saturation.

While the existence of cognitive sex differences in general has not been proven and meta-analysis shows that males and females are more similar than they are different (Hyde & Linn 1988; Hyde 2005; Miller & Halpern 2014), a female advantage has been found for specific tasks related to the production, memory and listening/analysis efforts in SI (Hyde & Linn 1988; Herlitz et al. 1997; Loonstra et al. 2001; Maitland et al. 2004; Hirnstein et al. 2014 for production; Trahan & Quintana 1990; Kramer et al. 1997; Kimura & Seal 2003; Gale et al. 2007 for memory; Hedges & Nowell 1995; Majeres 1999; Keith & Reynolds 2008; Aerts et al. 2013 for listening/analysis). Most of these studies have small or medium effect sizes, but they are sufficiently numerous for this predictor to be considered in interpreting studies. While there is a clear dearth of research on this topic in interpreting studies, these cognitive differences have inspired several researchers who have included sex as a predictor of Ear-voice span and disfluencies (Cecot 2001; Baes 2012; Defrancq 2013). Others have explored gender differences, i.e. differences in the interpreter's self-concept and role (Pöchhacker & Zwischenberger 2010; Magnifico &

Defrancq 2016, 2017). These studies have yielded significant results but have a very limited scope and small effect sizes. Accordingly, the aim of this research project is to consider the influence of a rarely analysed predictor (sex) on several variables during one specific task (simultaneous interpreting). The aim is neither to determine whether cognitive sex differences exist in general nor to generalise the findings to the rest of the population. Given the role played by memory in SI (Darò & Fabbro 1994; Hervais-Adelman et al. 2015), one could argue that the observed sex differences are likely to widen in a task where working memory is so intensely solicited. On the other hand, the intensive training and practice interpreters go through will potentially close the gap between the sexes.

This research project is based on a parallel acoustic aligned and time-tagged sub-corpus of the European Parliament Interpreting Corpus Ghent (EPICG). While corpora are traditionally used for descriptive studies, our hypotheses are exploratory. The cognitive processes involved in SI have been operationalised into four different research topics: the length of the Ear-Voice Span, i.e. the lag time between the source speaker and the interpreter, the rendition of numbers, the production of disfluencies and the use of extrapositions by interpreters working in Dutch and German, i.e. the fact of placing elements after the verb in order to shorten the amount of items to be retained in memory. The corpus consists of transcriptions of speeches and their interpretations recorded during plenary sessions of the European Parliament. The same data and languages have been used for all variables, except for the topic of extraposition, given that it only concerns interpreters working into SOV languages, i.e. German and Dutch. It was therefore decided to add 80 source and target German speeches to the corpus, but these German speeches could not be incorporated in the three other studies, due to time constraints. The data sample for the main sub-corpus consists of 180 source texts and 180 interpretations in six language pairs (from and into English, French and Dutch) for a total of 14 hours of interpreted speech and 108,245 interpreted words.

Simultaneous interpreting is a complex linguistic task that cannot be dissociated from the context in which it takes place (Diriker 2004; Duflou 2016). Therefore a study on the influence of sex without the inclusion of contextual variables would have been misleading. These predictors have been adapted to the topic under study but include among others the source and target languages, delivery rates and pausing patterns.

The hypotheses formulated for each topic come from similar assumptions: If women perform better at tasks linked to simultaneous interpreting, then they need to dedicate less energy to the concurrent cognitive processes involved in simultaneous interpreting and will therefore have more available cognitive resources and memory capacity. These assumptions lead to the four following research questions:

- Given that the length of the Ear-Voice Span reflects the cognitive processes involved in SI and is positively influenced by the intensity of the cognitive processes, will female interpreters have a shorter EVS than males?

- Will extraposition, being considered as a way of preventing the memory capacity to be overloaded, be used by female interpreters less often than males?
- Will disfluencies, being the consequence of cognitive resources being unavailable for a specific task, be produced by female interpreters less than by males?
- Omissions and errors are believed to be caused because of a lack of attentional resources required to perform the task. If numbers constitute a challenging task because of the additional cognitive resources they require, will female interpreters render more numbers correctly than male interpreters?

2. General implications

This dissertation helps fill two major gaps in interpreting research by including a rare variable, the sex of the interpreter, and a rarely applied method, the use of advanced corpus methods for exploratory research. Studies on sex differences can give interpreting researchers a new perspective on their work by encouraging them to consider sex as a significant variable. More generally, given the fact that the literature on cognitive sex differences in general still has not reached conclusive results and is often focused on isolated tasks, research on a complex and real task such as SI can help draw relevant conclusions. Moreover, while gender differences are not a focus for this study, they are also worth exploring and can be complemented by studies on cognitive sex differences. Together, they can help eliminate stereotypes or give potential explanations for yet unclear phenomena, such as the overwhelming predominance of females in the interpreting training and profession (Lim 2005; Ryan 2015; Hickey 2018).

In 1998, Shlesinger called for more corpus-based studies in interpreting in order to reinforce the empirical foundations of interpreting research. Indeed corpus data are naturalistic data produced in a real-life environment by professionals and therefore reflect the interpreting activity in a way experimental data cannot. Given the importance of the context and the communication in interpreting, it is crucial to have data from interpreters performing in a real setting. While plenary sessions at the European Parliament have specific features that hamper the generalisation of findings (short and read out interventions with a high delivery rate), the amount of data and the diversity of interpreters, source speakers, topics and languages allow for representative studies to be carried out. The institutionalised setting of the debates ensures the homogeneity of the data in regard to the interpreters' baseline qualification and working conditions. Moreover the metadata collected as part of the corpus compilation effort (Bernardini et al. 2017) guarantees that these interpretations do not consist of samples taken out of the context in which they occurred (Diriker 2004; Duflou 2016). Experimental research provides the most powerful design for testing causal hypotheses and allows for conditions to be controlled and is therefore often chosen for exploratory research. However, recruiting professional interpreters for experimental research is often tricky and most researchers heavily rely on students. While students' outputs are helpful and absolutely valid, the differences between them and that of professionals argue against applying findings from one group to the other. Moreover, contrary to translation corpora where texts have been through a deep reviewing process before being analysed, the interpreting product found in interpreting corpora is rendered with almost no revision or editing, given the time pressure, and therefore potentially reflects the cognitive processes involved.

Fortunately, the use of corpora in interpreting studies has gained popularity, notably thanks to the availability of online data and advanced tools for compiling corpora.

The corpus used for this study takes advantage of these new possibilities and can therefore help increase the use of corpora, as efforts have been deployed to guarantee the easy replicability of the corpus methods (for example the publication of a tutorial for Partitur) and the online availability of the corpus. One important advantage of this corpus is the existence of an oscillogram for source speech and interpretation, which facilitates the identification of ear-voice span, disfluencies and pauses (see Figure 1 in 3. Methodology). Moreover, a tailor-made script (see 3. Methodology and Appendix 1 for more details) has been developed in order to measure variables and their predictors automatically. The methodology and additional programs used to identify the sex and the uniqueness of interpreters in the corpus can also be replicated. Moreover, the unique methodologies developed to identify and operationalise the variables under study, as well as their potential predictors in the very specific context of simultaneous interpreting, can also be reproduced in subsequent studies. Most specifically, the segmentation of the speeches in 10-second segments which allows for the analysis of actual data instead of average values for a whole speech is particularly helpful, since most variables (EVS, delivery rate, disfluencies, etc.) can vary highly throughout a speech.

The results of the analysis on the specific variables chosen for each corpus study are also relevant for future research. Firstly, they are related to the cognitive processes involved in interpreting, notably memory management, and the process of dealing with cognitive load, topics that are increasingly popular. Indeed, the EVS is reputed to reflect the cognitive processes going on in the mind of the interpreters from the moment they hear a particular segment of the source speech until the point where an equivalent segment is uttered. Moreover, Gile (2008) considers EVS as a possible indicator of how the various efforts relate to each other. The EVS is also tightly linked to the interpreter's working memory capacity, as the longer the EVS, the more working memory capacity is required. Disfluencies can also be partly explained by the lack of mental resources required for the task. Extrapositions are also relevant for memory management, as comparing the point at which the verb is delivered in non-mediated language and in interpreting may reveal discrepancies that are significant for memory management in interpreters. Finally, numbers are seen as one of the most demanding tasks in interpreting and their omission or erroneous rendition is often linked to insufficient cognitive resources available. Demographic variables such as sex are interesting when looking at cognitive processes, because they consider the capacity dimension of memory. The human working memory has limited capacity, which prevents it from performing several tasks simultaneously at the same speed and the same level of efficiency as when the tasks are performed separately (Welford 1952; Broadbent 1958). Differences in the sex dimension could therefore be due to capacity differentials.

Secondly, innovative aspects have been included in each article and might inspire future research. Indeed, extrapositions in interpreters producing SOV structures have rarely been analysed, as studies in interpreting research mainly tackle this phenomenon from the perspective of the interpreters working from SOV structures. Moreover, the present study on the rendition of numbers is one of the sole corpus-based studies, since the vast majority of previous studies are based on an experimental framework. Furthermore, few studies have looked at the influence of variables on the interpreter's side when studying the Ear-Voice Span, such as disfluencies produced by the interpreter. Finally, most of the literature on disfluencies focuses on the tasks and their effect on the cognitive load and properties of interpreting, this study will also focus on the resources.

The current study also has potential implications for interpreting training. Studying sex differences is considered an important step to better understand what drives differential performance in professionals and in students and to help design effective training. While it is unlikely for sex-specific courses to be organized, sex differences would suggest that trainers need to take individual skills into account instead of offering one-size-fits-all solutions. More recently, corpus data have been put forward as a useful tool for trainees in knowledge acquisition and professionals in preparing the terminology for their assignments (Fantinuoli 2018). The variables under study are also relevant for training, notably because they often reflect strategies applied by professional interpreters and are taught in training courses. Firstly, the rendition of numbers is particularly challenging, professional interpreters apply different strategies to deal with them: shortening their EVS (Setton 1999; Jones 2002; Timarová et al. 2014); writing down the number (Setton 1999; Jones 2002; Mead 2015) and switching from intelligent hearing to literal hearing (Seleskovitch 1975; Pinochi 2009). As the difference in error rates between professional and student interpreters in the rendition of numbers shows, these strategies are not innate but are acquired with practice and experience and must therefore be taught in interpreting courses. The EVS is also a very tricky skill to acquire during interpreting training and our research gives insights on the predictors that influence its length, information that can be applied by trainers when guiding their students. Shorter EVS have been reported for experienced interpreters, which means that they are able to process the input faster, whether their shorter EVS is due to the use of strategies or to the automatisation of cognitive processes (Timarová et al. 2014). Disfluencies are also particularly frequent in students' renditions, and data from professionals can help students understand their cause and finally eliminate them. Finally, one of the most frequently mentioned contexts in which interpreters are said to need interpreting strategies is interpreting between languages with different word orders, in particular SOV-SVO. Extrapositions are an excellent strategy for interpreters working into German and Dutch. For interpreters interpreting from German and Dutch, strategies include waiting, chunking, anticipating, adopting a longer EVS, reformulating, compressing and omitting (Oléron and Nanpon 1965, 2002; Lederer 1981; Bevilacqua 2009; Seeber 2011). Given that strategies are very personal because they depend on the interpreter's strengths and weaknesses, demographic variables such as the interpreter's sex are particularly relevant.

3. Methodology

The sub-corpus of the EPICG used for this research project is still being compiled at the time of writing. Speeches are randomly selected on the European Parliament's website from plenary sessions between 2008 and 2014. For the transcriptions, the Valibel instructions (Bachy et al. 2007) have been adapted to facilitate the machine-readability of the texts. Besides the words spoken, other acoustic features are included in the transcriptions (e.g. disfluencies, self-repairs, false starts, filled and silent pauses). In its current shape, the corpus contains approximately 200,000 word tokens, with both source and target speech in English, French and Dutch. The data are compiled with the transcription suite EXMARaLDA (Schmidt 2012), which means that transcriptions and audio are aligned and that the software comes with a dedicated corpus compilation module and a concordancer. The detailed compilation, transcription and annotation process is described in Bernardini et al. (2018). Figure 1 presents the screenshot of an aligned transcription in EXMARaLDA, where the oscillograms of the source and target speeches are visible. Larger round grey areas on the oscillogram indicate that the speaker speaks, while quasi-flat lines represent silent pauses. Elongated sounds can be identified when the round grey areas start to extend along the flat line.

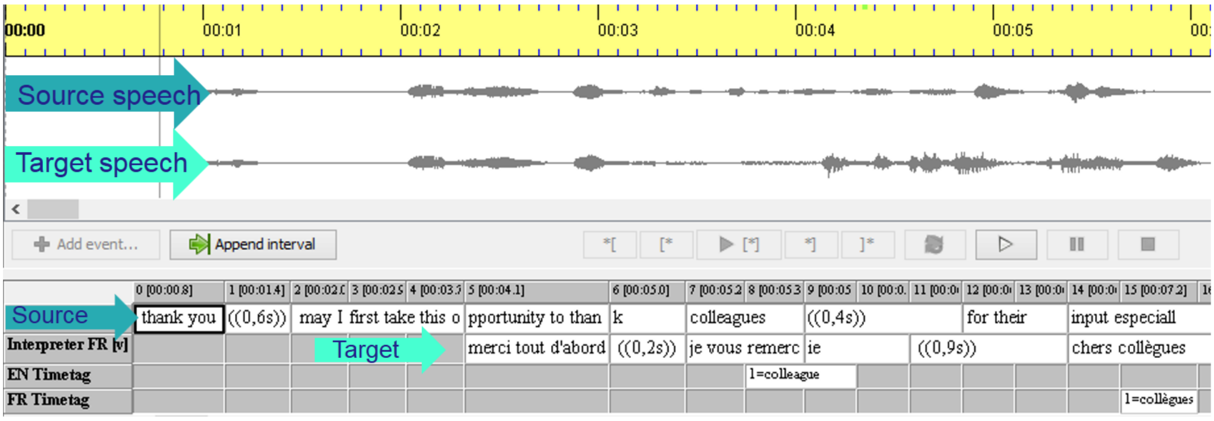


Figure 1. Screenshot from EPICG in EXMARaLDA

Data about the source speech and the interpretation included in each transcription are automatically extracted and sent to Excel files by a tailor-made script (i.e. a computer program written for this specific task). Some data are first calculated by the script (e.g. the delivery rate and the number of silent pauses) while others are simply copied from the metadata (e.g. the source speaker's name and interpreter's sex). The script coded in the programming language Tcl/Tk was specifically developed for this research project and is called "ProcessEXB" (EXB files are the files produced by EXMARaLDA). Thanks to the Appendix 1 which explains how to extract data from EXB files, it can easily be replicated. The Appendix also explains features of the script that were not used for this research project, such as the computation of data into range values.

For each combination of source and target language, a balanced set of 15 male and 15 female interpretations was aimed at. Given that the interpreters are anonymous, the sex of each interpreter was determined separately by three independent reviewers, as well as a speaker diarization software LIUM_SpkDiarization (Rouvier et al. 2013). The use and adaptation of this software for corpus data are explained in detail in Appendix 2 in order to enable replicability. Additional steps were undertaken in order to ensure a representative and diverse set of interpreters. Firstly, the corpus contains languages that are sufficiently common to guarantee that they are covered by a large number of interpreters. Secondly, speeches were randomly sampled from the European Parliament's website over a 6-year period in order to reduce the risk of having the same teams in the interpreting booths. LIUM_SpkDiarization was also used for the identification of identical interpreters and the results were rated as reliable by two human assessors.

When it comes to the identification of the four variables under study, methodologies have been adapted to our specific data. The Ear-Voice Span was measured by manually applying 10,864 pairs of time tags attached to equivalent lexical items to both the source text and target text. These time tags can be seen at the bottom of Figure 1 ("EN Timetag" and "FR Timetag"). The difference between the source and the target time tags is measured in centiseconds by the tailor-made script. The EVS length was measured for each pair of time tags, but also as an average per segment of 10 seconds and per interpretation. For our research, the following types of disfluencies were manually added to the transcription then measured by the script: lengthenings, filled pauses, false starts and silent pauses. These disfluencies are relatively easy to identify and were frequently observed in the corpus. For the rendition of numbers, a unique classification was adopted in order to represent the large diversity of renditions in a clear (for example the use of "substitution" instead of the more commonly used term "error" for categories) and systematic way that avoids overlaps (unlike several classifications from the previous literature). Moreover, in order to allow for studies with smaller samples of renditions, renditions were gathered into main and subcategories. Finally, the classification aims at making a distinction between errors and potential strategies. For example, a clear differentiation between implicit and explicit approximations is made, and omissions are considered as a specific category and are not as errors, as omissions are often associated with strategies. Moreover for a more detailed analysis of the intensity of errors, the percentage of elements being omitted or erroneous is added. Moreover, a specific EVS for numbers, i.e. the lag time between a number uttered by the speaker and the corresponding number in the interpretation, is also measured. More details can be found in the second corpus study. The methodology for the identification of extraposition is slightly more complex and requires additional information. Dutch and German are often referred to as SOV languages (Gerritsen & Stein 1992; Eisenberg 1994), because the verbal group can be placed at the end of a clause. In subordinate clauses, the main verb is situated near the end of the sentence, as in the example below:

> Es ist besonders wichtig, dass wir den Dialog mit Russland fortführen
> [it is particularly important that we the dialog with Russia continue]
> (EPICG_20080810_formal sitting1_Ingrid Betancourt_I_de)

This structure creates a *verbal brace* between the conjunction and the verb and the items placed in-between are called the *middle field*. Extrapositions happened when items from the middle field are placed after the verb, in order to reduce the number of items to be retained in memory. Our study focuses on subordinate clauses because they always consist of two poles, which is necessary to measure the length of the middle field. For each subordinate in the corpus, the number of words in the middle field was counted and labelled the "real middle field". In order to count the number extrapositions, it was necessary to identify items placed after the verb that theoretically could (or even should) be placed in the middle field. We therefore coined the concept of "theoretical middle field", comprising both the items found in the real middle field and those after the verb that could have been placed in the middle field. Extrapositions happen when the number of words in the theoretical field is bigger than the number of words in the real field.

Based on the overview of literature, pilot studies and metadata available in the corpus, several predictors besides the interpreter's sex have been included in the analyses. Some predictors are categorical and manually added to each transcription: source language and target language and source speech type. The other predictors are continuous and were measured thanks to the tailor-made script after having been manually identified in the transcription (delivery rate, source speaker/interpreter speaking time ratio, frequency and duration of pauses). Continuous variables are available in two conditions: averages normalized per minute, and actual values per 10 seconds (given that the speeches have previously been divided in segments of 10 seconds).

When analysing sex differences, Mann-Whitney U tests were used for the EVS, disfluencies and the use of extrapositions, while Chi-Square tests were performed for the rendition of numbers. In order to guarantee the independence of observations, the unit of analysis for Mann-Whitney U tests and Chi-Square tests performed in SPSS is the subject, i.e. the interpretation (rather than each EVS measurement or each disfluency for example). For the analysis of the influence of predictors, generalized linear mixed models were used for the EVS and the disfluencies, and logistic mixed models for the rendition of numbers (given that the dependent variable is categorical) as well as Chi-square tests for extrapositions. Mixed models were conducted on each segment of 10 seconds, rather than on single values per interpretation. Mixed models allow for the subject to be added as a random variable and therefore to control for idiosyncratic effects. As a consequence, the model “knows” that several observations come from the same interpretation. These models were performed with R with the following formula: “Response variable – predictor 1 + predictor 2 + predictor 3 + (1 | Interpretation)”.

4. Structure of the dissertation

This dissertation is divided into three main parts. The first part, introduction, covers the theoretical and methodological frameworks, as well as the complete introduction to this research project. The second part includes four corpus studies submitted as individual articles to peer-reviewed journals. Given that these articles are independent, redundancies are inevitable. The first chapter presents the study on the Ear-Voice Span published in *Perspectives*. This article being already published at the time of writing, an addendum has been inserted after the corpus study in order to include latest updates. The second chapter covers the rendition of numbers, an article submitted to *Interpreting*. The third Chapter concerns the analysis of disfluencies in a book chapter being reviewed at the time of writing for the series *Routledge Advances in Translation and Interpreting Studies*. Finally the use of extraposition is explained in the last chapter in an article accepted by *Meta*. The third and last part of this dissertation is the conclusion. Finally, two Appendixes are included. The first Appendix describes the tailor-made script designed for the analysis of the present data, while the second Appendix lays out the different steps necessary for the implementation of the speaker diarization software. These Appendixes are very detailed and require previous programming and statistical knowledge.

PART 2 : CORPUS STUDIES

Chapter 1: Predictors of Ear-Voice Span, a corpus-based study with special reference to sex

Abstract

This paper reports on a study on Ear-Voice Span (EVS) carried out on corpus data drawn from the European Parliament Interpreting Corpus Ghent, where sex is included as a predictor alongside several other variables. Ear-Voice Span is considered to be an indicator of cognitive processes in simultaneous interpreting and has therefore been selected to determine whether potential cognitive sex differences trigger different EVS patterns in men and women. Differences between men and women are reported in individual studies for tasks that are crucial to interpreting (Aerts, 2003; Kimura & Seal, 2003; Loonstra et al., 2001; Maitland et al., 2004 among others). However, meta-analyses tend to show that the reported cognitive differences between the sexes are exaggerated. This study uses corpus-based research methods to analyse the EVS of male and female interpreters in the European Parliament against the background of other known predictors of EVS. The data sample consists of 180 source texts and interpretations in six language pairs. The hypothesis was not confirmed as no sex differences were found. This research project helped identify relevant predictors of EVS: delivery rate, languages and interpreter's disfluencies.

Keywords: simultaneous interpreting; ear-voice span; corpus; sex differences

1. Introduction

As part of a broader research project on cognitive sex[1] differences in simultaneous interpreting (SI), this paper analyses the possible effect of sex, as well as other previously-investigated predictors, on ear-voice span (EVS). As such it is one of the very few studies on sex in interpreting. The scarcity of studies on sex differences in interpreting is surprising in many respects. Even though due caution is needed when it comes to research into cognitive differences between the sexes and meta-analyses tend to show that no clear difference exist, there seems to be evidence of sex differences in cognitive verbal tasks essential to SI (see Section 3). However this research project is not an attempt at proving the existence of cognitive sex differences in general, but mostly to explore hardly chartered territories.

Second, while the few studies on sex differences in SI and other interpreting modes are focused on specific populations of interpreters and can therefore not be generalized to all interpreters, they have shown interesting results that deserve to be investigated further. In small preliminary studies, Defrancq (2013) and Baes (2012) discover longer EVS for female interpreters at the European Parliament. In an experimental study on 11 professional interpreters, Cecot (2001) finds that women produce an average of 10.7% more filled pauses than men and men produce an average of 14.9% more unfilled pauses than women. Analyzing recorded courtroom proceedings, Mason (2008) finds that female consecutive interpreters in the courtroom tend to omit linguistic features that signal deference more than men, while male interpreters omit more politeness markers. Mason (2008) attributes these differences to cognitive, cultural and sociolinguistic variations between the sexes. While the above-mentioned findings are restricted in size and significance, they suggest that there might be a sex dimension to the way interpreters work that deserves a more systematic exploration. If sex appears to be a relevant factor in interpreters' performances, researchers would be well advised to consider its effect when studying interpreting.

Studies on cognitive sex difference in interpreting can also complement studies on gender differences (i.e. studies taking into account the interpreter's self-conception and role), either by giving potential explanations for the differences found or by suggesting that gender differences are mostly societal, cultural or educational and are not based on cognitive differences. While there is a surprising dearth of research focusing on gender

---

[1] For the purpose of this paper, differences between males and females will be described as sex differences. While the expression 'gender differences' is commonly used, it generally refers to an individual's self-conception and role within society. In fact gender differences studies tend to focus on communicative and linguistic differences (Chambers & Trudgill, 1998; Coates, 1993). The present study however focuses on the cognitive aspects and therefore, as for most studies described here, only takes the subjects' biological sex into account and makes no assumptions on their gender.

(Baer & Massardier-Kenney, 2016), some researchers have decided to tackle this topic. In corpus-based-studies, Magnifico & Defrancq (2016; 2017) have found that female interpreters at the European Parliament use more hedges and downtone fewer unmitigated face-threatening acts than male interpreters (with large effect sizes, respectively Φ=0.4 and 0.49). Analyzing 704 answers from professional interpreters to a survey on the perception of quality, Pöchhacker & Zwischenberger (2010) suggest that female interpreters rate others' performances more generously and value a lively intonation in interpreting more than male interpreters. However, while the differences are statistically significant, effect sizes are not mentioned. Angelelli (2004) found no gender difference when using gender as a category of analysis to study interpreter's attitudes toward their perceptions of roles and behaviors in practice.

Moreover, there is a clear imbalance between the sexes when it comes to the profession and to interpreter training around the world, female students outnumbering male students to a considerable extent (Lim, 2005; Ryan, 2015; Hickey, 2018). The reasons for the predominance of women in the profession remain unclear but Miller & Halpern (2014) suggest that individuals' relative cognitive strengths are important to their career and educational choices. Studies on cognitive sex differences could therefore help explain why few men opt for the interpreting career, while potentially eliminating stereotypes. Studies on sex differences in verbal tasks showing a female advantage might have an impact on men's decisions, as men might not choose interpreting based on their potential impression of a female predisposition for interpreting. This perception could also have an impact on their performance, given that people's performances appear weaker when they are told that the other sex performs better at the task (Spencer, Steele, & Quinn, 1999).

Finally, if sex differences exist, they could be taken into account when designing interpreter training. Indeed studying sex differences is considered an important step to better understand what drives differential performance in professionals and in students and help design effective training. Halpern et al. (2007) suggest that we can use this knowledge to teach female and male students ways to solve problems that correspond to their most efficient cognitive process to allow more flexibility in their problem solving and positively impact performance overall. While it is unlikely for sex-specific courses to be organized, sex differences would suggest that trainers need to take individual skills into account instead of offering one-size-fits-all solutions.

Our focus on EVS has several motivations: EVS is reputed to reflect the cognitive processes going on in the mind of the interpreter from the moment s/he hears a particular segment of the source speech until the point where an equivalent segment is produced. If these cognitive processes show a sex bias, EVS is likely to differ between groups of female and male interpreters. However, previous research has shown that EVS is determined by a plethora of other factors (Timarová et al., 2014). An analysis of EVS therefore cannot take sex in isolation as a predictor, but has to propose a model that includes sex among many other predictors. Such a model is presented in this study. Furthermore, EVS is a quantifiable property of interpreter performance, which partly explains its popularity in empirical interpreting studies: as a quantifiable variable it affords the kind of statistical analysis that is required to draw empirically sound conclusions on the studied predictors. In addition, EVS can be measured in a reproducible fashion, affording easy replication of research. Lastly, EVS is an important issue in interpreter training. In most handbooks on interpreting (Gillies 2013; Setton & Dawrant 2016), a whole section is devoted to EVS, advising students to maintain EVS at a level allowing them to unravel the segment's meaning, shortening it to tackle specific problems such as numbers or names. If EVS displays sex patterns, it would of course be advisable to bear these in mind while teaching students to maintain an appropriate EVS.

The research reported in this article is based on a parallel acoustic aligned and time-tagged sub-corpus of the European Parliament Interpreting Corpus Ghent (EPICG) that is currently being compiled at Ghent University. It consists of transcriptions of speeches and their interpretations recorded during plenary sessions of the European Parliament. Corpora are designed to reflect the variety of linguistic phenomena and paralinguistic conditions that are representative of language in use. Even though the plenary sessions at the European Parliament may not be an entirely adequate response to the shortcomings of experimental data, as they are generally very short (1 to 6 minutes), the amount and diversity of the data allow for a representative study to be conducted. Moreover, the institutionalized setting of the debates ensure that the data in regard to the interpreter's working conditions are fairly homogeneous. The accreditation tests interpreters must take to be allowed to carry out interpreting assignments in the European institutions ensure a baseline interpreting quality in the data.

In Section 2, the concept of ear-voice span is presented as an indicator of the cognitive processes involved in simultaneous interpreting. Section 3 covers the literature on sex differences for several cognitive skills relevant to simultaneous interpreting. The research question and hypotheses are developed in Section 4. Section 5 presents the corpus and the methodology used to measure the ear-voice span and the predictors. The results of the statistical analyses are reported in Section 6. Finally, Section 7 presents the discussion of these results.

2. Cognitive processes in simultaneous interpreting and Ear-Voice Span

The concept of ear-voice span is popular among researchers in interpreting studies. The reason behind this interest might be that EVS, while relatively easy to measure, has the potential to unveil some of the cognitive processes underlying simultaneous interpreting (Lee, 2002; Timarová et al., 2011). Simultaneous interpreting is a complex task involving several cognitive processes: speech comprehension and production, memory, attention/resource allocation and coordination (Pöchhacker, 2015). In order to explain variations in interpreting performance, Gile (1995; 2018) puts forward an Effort Model for simultaneous interpreting that includes four processes, called efforts, which compete for limited attentional resources (also called processing capacity). The first effort is the listening and analysis effort which includes the detection and identification of stimuli and the assignment of a meaning to what is heard. Secondly, the short-term memory effort is presented as a storage mechanism where information is temporarily kept before further processing takes place (Liu, 2008). The production effort represents the planning and production of the speech in the target language, as well as self-monitoring. Finally, the coordination effort accounts for the management of attentional resources to the three other efforts, given that these resources are believed to be limited. After conducting neuroimaging studies during simultaneous interpreting, Hervais-Adelman, Moser-Mercer, & Golestani (2015: 1) confirm that simultaneous interpretation 'places extreme demands on the cognitive control of language and on verbal working memory and attention.' The constraint of simultaneity and the limited capacity make interpreting particularly challenging.

Every individual cognitive task, even the simple task of repeating a word, requires a certain amount of time for completion and the duration of that amount of time can be a good indicator of the processes involved. Mizuno (2017) draws a parallel between interpreting studies and cognitive psychology where reading difficulty is often measured by latency (the time elapsed for the completion of a task). Comparing interpreting with tasks that do not involve a translation process, such as the verbatim repetition of words and sentences in the same language, studies find longer EVS for interpreting (Christoffels & De Groot, 2004; Gerver, 1969; Treisman, 1965). These studies show that lag time (in this case the ear-voice span) reflects cognitive processing and that the more complex the task, the longer the lag (Timarová, 2015).

Moreover, Gile (2008) considers EVS as a possible indicator of how the various efforts (or cognitive processes) in SI relate to each other. Indeed, a long EVS might mean that the interpreter prioritizes the listening effort over the production effort, while a short EVS potentially means that the interpreter is saving working memory capacity. Gile (1995; 2018) also points out that simultaneous interpreters tend to work close to cognitive saturation (referred to as the tightrope hypothesis), which means that omissions or errors can be caused because attentional resources required to perform adequately were not available for a particular comprehension, memory storage or retrieval or production task at a time when they were needed. Kade & Cartellieri (1971) estimate that the optimal moment for the interpreter to start speaking is immediately after all syntactic and semantic ambiguities in the unit have been resolved and Goldman-Eisler (1972) found that interpreters usually wait for the predicate of a sentence before they start interpreting it. While interpreters decide when to start speaking and therefore determine the amount of time dedicated to a task (or the EVS), they are also limited by their working memory capacity (Gile, 1999). In fact interpreters' working memory capacity does not always allow them to wait long enough and, by fear of overloading their memory, they are often tempted (or forced) to start interpreting before knowing exactly what is meant running the risk of misinterpreting the speaker. Having a long EVS has stylistic advantages: the interpreter has more time to reformulate and avoids literal translation. But having a long EVS can also have drawbacks as research indicates that a lag time longer than 4 seconds leads to reduced accuracy (Lee, 2002; Timarová et al., 2014). The interpreter therefore needs to find a compromise between the length of input required for understanding the speaker and the limits of working memory capacity. EVS is thus tightly linked to the interpreter's working memory capacity and the longer the EVS, the more working memory capacity is required.

With regard to variables influencing EVS, Timarová (2015) argues that cognitive (and memory) limitations can only explain the minimum lag (determined by the speed of processing) and the maximum lag (determined by memory capacity) an interpreter can keep and that the variation between these limits may be explained by other variables. Most studies on EVS analyze variables linked to the source speech. Goldman-Eisler (1972) found a longer ear-voice span for interpreters from German to English than from English to French or French to English, which he attributes to the fact that the German verb is often at the end of a sentence (SOV order), therefore the interpreter from German may have to wait for the verb in the input, causing a lengthening of the ear-voice span. The speaker's delivery rate and speech/pause ratio (total speaking time divided by the total duration of silent pauses) determine the rate at which information has to be processed by the interpreter and could therefore influence the EVS. De Groot (1997) argues that high input rate means the time span over which the words are presented to the interpreter in the input is relatively short, therefore the processing of information is more difficult. In an experimental study, Gerver (1969) discovered that the EVS increases progressively if the input rate increases and showed that this increase in EVS length is the result of interpreters' failure to speed up their delivery rate to the same extent as the speaker's. In contrast, Lee (2002) studied Korean interpreters and found that high input rate reduced EVS. This might be due to the fact that a lot of information in a short amount of time (implied by a high delivery rate) could overload the memory capacity and encourage the interpreter to shorten the EVS in order to save working memory capacity. Lee found that

EVS increased as a result of shorter pauses between sentences in the source speech. The shorter the pauses between sentences, the more the interpreter must listen and speak at the same time, which makes the task more cognitively demanding and, as a results, increases the EVS. He also found that Korean interpreters began speaking earlier (and therefore had a shorter EVS) after a segment in which they were silent and only listened to the speaker, than after a segment in which they had to speak and listen at the same time, once again showing that short EVS is linked to a reduced cognitive effort.

Source text type also appears to influence EVS: Barik (1973) found smaller EVS in interpreting an impromptu speech as compared to a pre-written speech. Written speeches are associated with a higher speech/pause ratio. Lee found that the EVS is longer when the original speaker uses longer sentences. The syntactic complexity of the source speech can mean that the interpreter needs to wait longer to process the input (as the interpreter must also process the syntactic information) before starting to speak and therefore increase the EVS. Some parts of a speech that are non-contextual and highly informative, like lists, names and numbers are more demanding than others and seem to require a shorter EVS (Kader & Seubert, 2014). Setton & Dawrant (2016) indeed recommend interpreters in training to adopt a shorter EVS when dealing with non-contextual items. Finally, Díaz Galaz (2011) found longer EVS for difficult source speeches (text difficulty being determined by the presence of terminology, syntactic complexity and the presence of non-redundant items, such as proper names and figures).

Fewer studies look at the variables on the interpreters' side. Lee (2002) found that the EVS increases when the interpreter uses longer sentences, which can be explained by the increased effort of producing long sentences. Lee (2002) also found that the EVS obviously decreases when the interpreter produces more syllables and speaks more than the original speaker. Moreover, interpreters with long EVS have higher words per minute rates and speech/pause ratios, which might imply that when the EVS gets longer, interpreters speak faster in an attempt to reduce their EVS. Díaz-Galaz, Padilla, & Bajo (2015) report smaller EVS following advanced preparation by the interpreter and suggests that the short EVS is linked to facilitation of cognitive processes through preparation. Timarová et al. (2014) report shorter EVS for interpreters with more experience and suggests that they are therefore able to process the input faster, whether their shorter EVS is due to the use of strategies or to the automatisation of some cognitive processes. Lamberger-Felber (2001) found longer EVS for simultaneous interpreting with text compared with simultaneous interpreting without text, which could be explained by the fact that interpreters working with text suffer less from memory restrictions and can afford to have a longer EVS. Several sources stress that individual interpreters' EVS varies strongly during the interpretation of one speech and that the EVS of several interpreters assigned the same speech is also very variable (Lamberger-Felber, 2003; Timarová, et al., 2011). This variation in EVS seems to indicate that the interpreter's preference has an influence on the EVS because interpreters might use different cognitive processes and strategies depending on their strengths and weaknesses.

To summarize, the length of EVS seems to be influenced by factors related to the interpreter and to the source speech which, among other things, determines the complexity of the task. Variables on the source text and interpreter side include the language, the delivery rate, the speech/pause ratio, the syntactic complexity and the original/interpreter speech ratio. On the source side only, the following variables have been identified as potentially influencing the EVS: the type of delivery, the lexical density and the type of information. Other variables only concern the interpreter: cognitive and memory limitations, personal preferences and strategies, distribution of resources between the several cognitive efforts, preparation, experience. This research project will include some of the above mentioned variables (see Predictors in Section 5).

3. Sex differences in cognitive skills

The topic of sex differences in cognitive skills is complex and needs to be handled with care and nuance. In recent years, many researchers have grown skeptical about claims of cognitive sex differences and meta-analyses tend to indicate that males and females are much more similar than they are different and that sex differences are often exaggerated (Hyde, 2005; Hyde & Linn, 1988; Miller & Halpern, 2014). When it comes to sex differences in verbal abilities more specifically, Hyde & Linn's meta-analysis (1988) shows that the magnitude of sex differences is currently so small that it can effectively be considered to be zero. While meta-analyses are essential for determining whether sex differences exist as a whole or in general cognitive abilities, Hyde & Linn also suggest that we study sex differences in abilities more precisely and 'move away from the old model of intellect that specified only three rather general cognitive abilities - verbal ability, mathematical ability, and spatial ability' (p 33). Accordingly, the aim of this research project is neither to determine whether cognitive sex differences exist in general and outside of interpreting nor to generalize the findings to the rest of the population. This research project aims at exploring the influence of a rarely analyzed predictor (sex) on one variable (EVS) during one specific task (simultaneous interpreting).

Hyde & Linn's meta-analysis on verbal abilities includes various and diverse tasks, among which several are not specifically relevant for SI: spelling, reading, writing and vocabulary for example, and for which no substantial sex differences are found. Hyde & Linn do recognize that there is one possible exception for which females score higher than males and which is highly relevant to interpreting, namely speech production, with an effect size of 0.33. However, several statistical sources would consider this effect size to be small (Ferguson, 2009; Mellinger and Hanson, 2017) and other tasks related to verbal abilities show no sex difference (vocabulary with an effect size of 0.02 and anagrams with an effect size of 0.22) or give higher values for men (analogies with an effect size of -0.16). Besides the meta-analyses, which fail to confirm the existence of sex differences, several individual studies claim to find female advantage in a series of abilities related to simultaneous interpreting. Unfortunately, most studies do not mention effect sizes (or do not give sufficient information for these to be calculated by the reader). Moreover, when effect sizes are mentioned, they are generally small, ranging from 0.3 to 0.6. Several studies conclude that women have greater verbal fluency, i.e. the ability to retrieve specific information within restricted search and time parameters, for example the ability to generate words beginning with a single letter in one minute (Herlitz, Nilsson, & Bäckman, 1997; Loonstra, Tarlow, & Sellers, 2001; Maitland et al., 2004). Verbal fluency is believed to play a key role in interpreting (Stavrakaki, Megari, Kosmidis, Apostolidou, & Takou, 2012). A female advantage in generating synonyms has also been found (Hines, 1990). For the listening and analysis effort, sex seems to influence the phonological input processes activated when the human brain perceives a spoken word and tries to identify it. Aerts (2013) found that women display a larger sensitivity to the phonemic contrasts during auditory phoneme discrimination and showed more differentiation in real word-pseudoword dissociation. Studies have found faster processing speed (the speed at which a person can understand and react to the information they receive) in women (Keith, Reynolds, Patel, & Ridley, 2008) and a female advantage in both prelexical and lexical processing was found (Majeres, 1999), as well as for perceptual speed (the ability to compare or recognize items) (Born, Bleichrodt, & van der Flier, 1987; Hedges & Nowell, 1995). Finally, evidence for a female advantage in episodic and some aspects of semantic memory has been found (Herlitz, Airaksinen, & Nordström, 1999; Kramer, Delis, Kaplan, O'Donnell, & Prifitera 1997; Maitland, Herlitz, Nyberg, Bächman, & Nilsson, 2004). Several studies have also shown that women perform better than men on immediate and delayed free recall and on recognition tasks with verbal and visual components (Kimura & Seal, 2003; Trahan & Quintana, 1990). However, Harness, Jacot, Scherf, White, & Warnick (2008) report higher scores for males in a study on recall combined with a distraction task carried out on students. Other studies report that females outperform males in the Rey Auditory Verbal Learning Test (free recall of two lists of nouns, which aims at evaluating short-term auditory-verbal memory, retroactive, and proactive interference, retention of information among others) and the Verbal Paired Associates test (immediate and delayed recall of word pairs, aimed at evaluating explicit episodic memory performance) (Bolla-Wilson & Bleecker, 1986; Gale, Baxter, Connor, Herring, & Comer 2007).

4. Research question and hypotheses

Simultaneous interpreting involves several cognitive processes and the ear-voice span is considered a good indicator of these processes. EVS has therefore been chosen as the dependent variable in this research project. While several predictors influencing the EVS have been explored in interpreting studies, one variable, the sex of the interpreter, has been neglected, and has therefore been included in this research project.

Sex differences have been observed for several of the cognitive abilities involved in SI (production, analysis and memory efforts) and most studies show a female advantage. Therefore, our main research question is: do these sex differences in cognitive skills influence the EVS? Because of the high requirements in quality and skills for the interpreters working at the plenary sittings of the European Parliament, we assume that all interpreters studied in the corpus are professional interpreters who, in most cases, render an interpretation of the original speech that fulfills the requirements of their profession. In other words, we assume a minimum level of quality for male and female interpreters that was established by their accreditation test. The aim of this research project is therefore to determine whether male and female interpreters use different EVS to fulfil the minimum level quality required for their job.

If we assume that

- women need to dedicate fewer cognitive resources to the interpreting task than men because of the female advantage in the production, analysis and memory efforts;
- the length of the EVS is positively influenced by the intensity of the cognitive processes (and more specifically the working memory capacity) (Lee, 2002; Timarová et al., 2011);
- interpreters dedicate almost all their processing capacity to the task they perform (tightrope hypothesis, Gile, 2008);

we can expect men to have a longer EVS than women.

Given the numerous predictors of EVS identified in the literature, this study cannot analyze sex as the sole predictor of EVS. Therefore, 15 additional potential predictors of EVS have been analyzed. These predictors have been chosen because of their relevance in the literature and of results of pilot studies on the corpus, we therefore expect them to have a strong influence on EVS.

5. Materials and methods

5.1 Corpus-based interpreting studies

Most studies on EVS are conducted in the framework of an experimental research design. Experimental research offers the advantage of controlled conditions, but Shlesinger (1998) argued that interpreting corpora could reinforce the empirical foundations of interpreting research. Indeed corpus data are naturalistic data produced in a real-life environment by professionals and therefore reflect the interpreting activity in a way experimental data cannot. Moreover when corpora are available online, they allow researchers to reproduce research results and replicate studies.

Use of corpora is becoming increasingly popular in interpreting studies. New technologies have offered solutions for the time-consuming compilation, transcription and analysis of interpreting data with tools such as EXMARaLDA (Schmidt & Wörner, 2009), Praat (Boersma, 2001) and SpeechIndexer (Szakos & Glavitsch, 2007) among others. Moreover, for a number of years, the plenary sittings and some of the committee sittings of the European Parliament can be downloaded from the website of the European Parliament[2]. Corpus-based studies are sometimes criticized because they consist of samples taken out of the context in which they occurred (Diriker, 2004; Duflou, 2016). Therefore the inclusion of metadata is necessary to provide more context and it is important to add as much contextual information as possible when building a corpus. The European Parliament website also gives access to several metadata about the speaker (political group and function, age, sex) and the speech (topic, time of the day, delivery type). The first consistently compiled simultaneous interpreting corpus to have become publicly accessible is the European Parliament Interpreting Corpus (EPIC), compiled at the University of Bologna from recordings of European Parliament sessions (Bendazzoli & Sandrelli, 2005). Several other universities have also started compiling their own corpora and have developed useful tools and guidelines for corpus-based interpreting studies : CoSi for consecutive and simultaneous interpreting (House, Meyer, & Schmidt, 2012) and DiK for dialog interpreting in public service settings (Bührig, Kliche, Pawlack, & Meyer, 2012) at the University of Hamburg, EPICG (European Parliament Interpreting Corpus Ghent) at Ghent University (Bernardini, Ferraresi, Russo, Collard, & Defrancq, 2018), and others at the universities of Rome, Trieste, Posnan, Louvain-la-Neuve and Saarbrücken.

---

[2] http://www.europarl.europa.eu/ep-live/de/plenary/

5.2 The corpus used

Like EPIC Bologna, the parallel acoustic aligned corpus of EPIC Ghent consists of transcriptions of speeches and their interpretations recorded during plenary sessions of the European Parliament. Source and target texts were transcribed according to the Valibel instructions (Bachy et al., 2007) with some adaptations to facilitate the machine-readability of the transcription. Source and target texts are acoustically aligned on the basis of pauses with the transcription tool EXMARaLDA Partitur-Editor. More information on the compilation process, including transcription conventions and annotations can be found in Bernardini et al. (2018). For this study, 30 source speeches of each of the six following combinations were randomly selected: English-French, French-English, English-Dutch, Dutch-English, Dutch-French and French-Dutch. For each target language, a balanced set of 15 male and 15 female interpretations was aimed at3. In total, the corpus comprises of more than 14 hours of interpreted speech and a total word count of 108,245 interpreted words. As the authors did not have access to the interpreters' identities, two methodological challenges arose while selecting the speeches and their interpretations: on the one hand, sex had to be determined on the basis of the properties of the recorded voices; on the other hand, a sufficiently varied sample of interpreters had to be included to avoid a possible bias resulting from a sample dominated by a limited set of interpreters only. The sex of each interpreter was determined separately by both authors and an independent reviewer. According to Lass & Puffenberg (1971), human identification is a reliable method, as listeners are able, even on the basis of vowels pronounced in isolation, to identify speaker sex with an accuracy of over 95%. The process yielded an inter-rater agreement of 99.4%, with the three assessors diverging on only one interpretation. It was concluded that the disagreement came from an human encoding mistake and the three assessors finally agreed on all interpretations. In order to complement the human identification process, a speaker diarization software (LIUM_SpkDiarization, Rouvier et al., 2013) was used on the corpus to identify the interpreters' sex. After several necessary modifications to the software and the corpus's audio data, the software agreed on all identified sexes except for 8, reaching an agreement of 95.6%. However two human assessors disagreed with the 8 sexes identified by the software and agreed with the human identification. This discordance can be explained by the fact that the software is optimized for radio and TV shows and the same level of performance cannot be expected for other types of recordings. Moreover, the corpus's audio data are complex (several

---

[3] According to the estimations of the Organisation Intersex International Europe, there is a probability that one percent of the interpreters in the corpus are intersex. Unfortunately there is no possibility of knowing whether interpreters in the corpus are intersex. While the authors are conscious of this possibility, they consider that the size of the corpus means that this low probability does not obviate much of the discussion.

speakers take the floor simultaneously and the quality is not always optimal) and the diversity of languages and speakers is high. Considering the high human-machine agreement and the fact that the 8 sexes for which humans and software disagreed contained specific challenges, in particular for the software (several interference with speakers from the floor), and given that human inter-rater agreement in these 8 cases was 100% , it was decided to keep these data and trust the human identification.

To tackle the risk of an unrepresentative set of interpreters, a number of steps were taken. First, the languages chosen for this study (English, French and Dutch) are sufficiently common to guarantee that they are covered by a high number of interpreters. Second, speeches were manually sampled from the European Parliament's website with a specific aim to reduce the risk of sampling multiple interpretations by the same interpreter. Therefore, the period from which the recordings were taken stretches over 6 years and sessions were picked randomly to collect the interpretations. In the final dataset, 93 interpretations are drawn from 21 different dates in 2008, 47 interpretations from 17 different dates in 2009, 14 from 11 different dates in 2010, 9 from 9 different dates in 2011 and 16 from 4 different dates in 2013 and 1 in 2014).

While humans seem to be more successful than the software at identifying sex in this corpus, the identification of identical speakers for such a diversified corpus appears to be very unreliable when done by human listeners. Indeed an attempt was made by humans to identify similar interpreters in the corpus but the amount of different interpreters made this task almost impossible. Therefore, LIUM_SpkDiarization was used for the identification of identical interpreters and the results were considered reliable by two human assessors. However, given the difficult for human listeners to identify the interpreters, no clear human-machine agreement could be determined. Therefore table 1 presents the results of the LIUM_SpkDiarization's analysis only. They show that the maximum number of interpretations by one and the same interpreter is 3, which is 10% of the interpretations included for a particular booth.

Table 1. Identification of identical interpreters by LIUM_SpkDiarization

Sex	Language	Number of interpreters identified		Total number of unique interpreters
		Twice	Three times	
Females	French	3	0	27
	English	3	1	25
	Dutch	4	0	26
Males	French	3	0	27
	English	2	2	24
	Dutch	2	0	28

While theoretically having several identical interpreters in the data set could violate the assumption of independence of observations within each group (males and females), the authors believe that the independence is not compromised for several reasons. First, given the large size of the corpus and the limited number of potentially identical interpreters, the dependence of observations would only concern a small number of interpretations and the final corpus can be considered as representing a sufficiently diverse set of interpreters. Second, each interpretation, even if performed by the same interpreter, is unique and performed in different conditions (conditions that have an influence on the EVS). Third, data are not aggregated at the individual level, but at the group level (males and females, where the independence of observations is guaranteed, as the same interpreter cannot be included in both the male and the female group).

5.3 EVS measurement

The ear-voice span is the delay between the speaker's input and the interpreter's output (Timarová, 2015). Regarding the EVS length, the results of most studies more or less coincide on an average of 3 seconds (Oléron & Nanpon, 1965) or 4 to 5 words (Treisman, 1965). Even if most studies agree on an average EVS, the methodologies used to measure it often differ. The tendency is mostly to measure EVS by identifying semantically equivalent lexical items in the source and target texts but there are differences as to the unit of measurement used. Some studies measure EVS in words or other linguistic constituents (Donato, 2003; Gerver, 1969; Goldman-Eisler, 1972, Treisman, 1965). The most common practice however is to measure it in centiseconds or milliseconds, as this method seems to be less influenced by diverging word structure across languages (Setton, 1999). With the growing use of technology in the transcription and processing of data, measuring EVS has become easier. There remains some disagreement on where exactly the measurement should start and end. Christoffels & De Groot (2004) and Ono, Tohyama, & Matsubara (2008) specify that the end of EVS coincides with the onset of the target language item but disagree on whether measurement should start from the onset (Christoffels & De Groot, 2004) or the end (Ono et al., 2008) of the source language item. Researchers also choose different intervals between items used as reference points: Gerver (1969) measures EVS every fifth word of the source text, while Barik (1973) uses one source language item every five seconds. Christoffels & De Groot (2004) place reference points on three words evenly distributed across every other sentence and measure the average of each set of three words. Ono et al. (2008) mark all content words in the source text. Timarová et al. (2011) compared three methods (Barik, 1973; Lee, 2002; Treisman, 1965) and indicated that their mean EVS were not significantly different. This being said, the EVS length is not a fixed value and mean or median values for a whole interpretation might not be representative of its complexity. For this study, it is important to note that the decisions on EVS measurement were made with the study's aim in mind, which is to compare EVS lengths in different conditions and not to determine a general average value for EVS. The main criteria were therefore homogeneity and comparability.

In accordance with most studies, the EVS for this study was measured with time tags attached to equivalent lexical items applied to both the source text and target text. Similarly to other studies (Christoffels & De Groot, 2004, Timarová et al, 2014), the scale used for EVS in this study is centiseconds, as the transcription tool used for the corpus (EXMARaLDA Partitur-Editor) does not allow for precise identification of onset or end of lexical items at the scale of milliseconds. A pair of time tags was manually and randomly added every 10 items uttered by the interpreter. The interpreter's output was chosen as the reference text simply because the EVS cannot be measured if items have not been interpreted. For this research project, items chosen as time tags are of any grammatical category (substantives, nouns, verbs, etc.). However, in order to reduce the influence of language structures, the two items chosen for a pair of time tags are of the same grammatical category. For the same reason and to ensure homogeneity, the items with the shortest EVS are chosen whenever different word orders and, therefore, different EVS measurements are possible within the same word group (e.g. for the two equivalent expressions ‘Union européenne’ and 'European Union’, the time tags are added to ‘européenne’ and ‘European’ and not to ‘Union’ and ‘Union’). For source texts and target texts considered as a whole, a total of 10864 pairs of time tags have been applied by the transcribers (and verified by at least two reviewers).

After measuring and analysing the EVS lengths in different conditions (from the onset, middle or offset of the source and target items) with a tailor-made script, it was noted that the EVS lengths were similar across conditions for all languages, notably because the average number of characters per word is very similar across languages (4 to 5 characters per word) and therefore does not influence the EVS length. The authors therefore decided to present results for the first condition (from the onset of the source item to the onset of the target item) in order to allow comparability with other studies on the topic. In order to measure the EVS and observe the influence of predictors in a short period of time (instead of looking at the average values of EVS for a whole speech), the speeches have been split in 10-second segments and the average EVS per 10 seconds has been analysed.

The distributions of EVS being rightly skewed (even after log-transformations), the non-parametric one-tailed Mann-Whitney U-test was chosen to check for potential sex differences in EVS. The assumptions for the test are fulfilled: the dependent variable (EVS) is continuous, the observations are independent (an interpreter cannot be represented both in the male and the female group) and the distributions of both groups (males and females) have a similar shape.

5.4 Predictors

Based on previous literature, pilot studies and metadata available in the corpus, sixteen predictors have been identified as potentially relevant for the EVS: interpreter's sex, source language, target language, source speaker's and interpreter's delivery rate (measured in number of words per 10 seconds), speech/pause ratio (a pause is a silence of more than 0.2 second), average duration of silent pauses, length of segments between pauses (measured in number of words between silent pauses), the source speaker's/interpreter's speech ratio (total speaking time of the original speaker divided by the interpreter's total speaking time), the source speaker's delivery type (impromptu, mixed or read, conforming to the encoding in Bernardini et al., 2008), as well as interpreter's disfluency (number of filled pauses, words with elongated pronunciations, also called lengthenings, and false starts per 10 seconds). Other predictors were mentioned in the introduction but could not be included in this analysis because they are not available. Indeed since the interpreters are anonymous, the authors have no information about their personal preferences and strategies, their level of preparation or their experience. Similarly, the syntactic complexity and lexical density of the data are not measured in our corpus.

Some predictors are manually added to each transcription (language, sex, type of delivery) while others are automatically measured by a tailor-made script after having been manually identified in the text. Most predictors are averages measured on segments of 10 seconds, instead of averages measured for a whole speech, as we consider that these data can vary highly throughout a speech both for the speaker and the interpreter. The type of delivery (impromptu, mixed or read) was determined by watching the video of the source speech. When the speaker reads from a document while speaking, the type is 'read', when the speaker alternatively reads off a document and speaks without looking at the document, the type is considered as 'mixed'. When the speaker does not have a document, the type is 'impromptu'. However, some video data were not available and the information about delivery type is sometimes incomplete.

After the Mann-Whitney U test performed to identify potential sex differences, an exploratory multiple regression analysis (forward hierarchical regression) was conducted with all predictors in order to determine their individual significance. The skewness of EVS does not prevent a regression to be carried out as only the residuals need to be normally distributed. The results are presented through the β-coefficient, which indicates the individual contribution of each predictor to the regression model. The standardized versions of the β-coefficients are also mentioned as they are easier to compare (they are not dependent on the units of measurement of the variables), as well as the R squared (R2), which is a goodness-of-fit measure.

6. Results

6.1 Descriptive statistics

The mean value for the EVS for the 10864 EVS measurements is 3.03 seconds (M=2.69 seconds and SD=1.64 seconds) which is in line with most findings in the literature (the results of most studies coincide on an average EVS of 3 seconds). As figure 1 shows, data for the EVS are not normally distributed and rightly skewed, which is typical of response latency tasks.

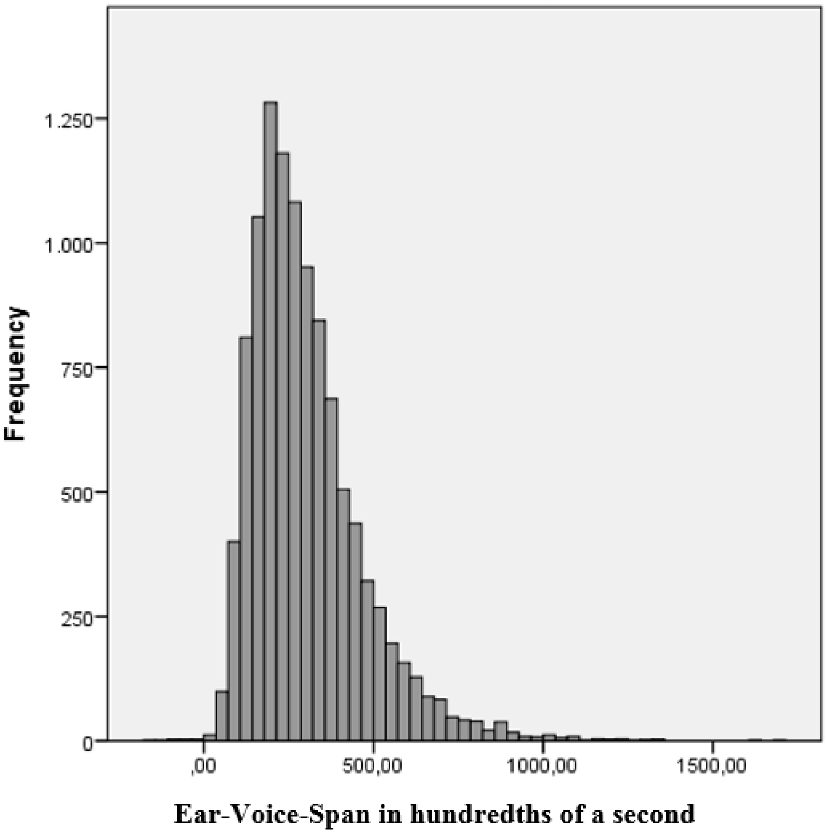


Figure 1. Histogram of EVS

Many outliers are present in the data. After verification, these outliers are not due to errors in the dataset but are simply linked to the nature of the data and have therefore not been removed.

EVS lengths range between -156 and 1687 centiseconds (-1.56 and 16.87 seconds). A negative EVS implies that the interpreter is ahead of the speaker and is (in most cases) anticipating what the speaker is about to say. As a result, the tag for the item concerned occurs first in the target language and then in the source language. Where EVS is 0 second, the source and target language items coincide. In this case, the interpreter is not ahead of the speaker but has evidently not had time to hear the source language input and must therefore have anticipated it. However, negative EVS only accounts for uttered anticipation, but not for extra linguistic anticipation where the interpreter knows what the speaker is going to say but decides not to utter it before the speaker. In total, there are only 11 negative EVS (8 for male interpreters). Therefore uttered anticipations account for about 0.1% of the data. Instances of EVS shorter than 2 seconds account for 29% of the data while EVS longer than 4 seconds account for 21% of the data. EVS between 2 and 4 seconds therefore account for 50% of data. As mentioned in the methodology, the average EVS per 10 seconds was also measured and the same analyses were conducted. The results are very similar to the data presented for each measurement of EVS and were therefore not included.

The descriptive statistics regarding the predictors are shown in table 2 and are averages per 10 seconds. It appears that the original speaker has a higher delivery rate than the interpreter. However, the interpreter has a higher speech/pause ratio and a higher average sentence length. The most frequent type of disfluencies for the interpreter is filled pauses.

Table 2. Descriptive statistics for predictors

Predictor	Mean	Median	SD
Interpreter's delivery rate (in words per 10s)	23.63	23.61	5.98
Original speaker's delivery rate (in words per 10s)	26.07	25.76	5.14
Original speaker/interpreter speaking time ratio	1.04	1	0.42
Interpreter's number of filled pauses	1.06	0.98	1.23
Interpreter's number of lengthening	0.54	0	0.85
Interpreter number of false starts	0.25	0	0.54
Interpreter speech/pause ratio per 10s	7.14	5.67	6.07
Original speaker speech/pause ratio per 10s	5.82	4.97	3.59
Interpreter's average sentence length per 10s (in words)	6.2	5.33	3.57
Original speaker's average sentence length per 10s (in words)	5.62	5.03	2.68

6.2 Sex differences in ear-voice span

The mean value for EVS among females interpreters is 3.01 seconds (M=2.67 seconds with 5420 EVS measurements) while the EVS for male interpreters has a mean value of 3.05 seconds (M=2.71 seconds for 5444 EVS measurements). Both distributions are rightly skewed (see figures 3 and 4) and have a similar shape.

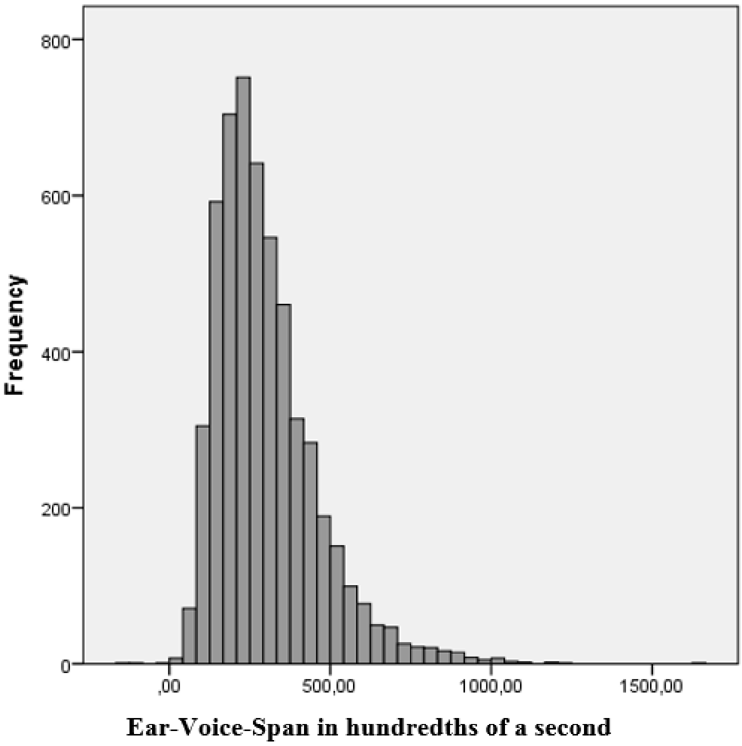


Figure 2. Histogram EVS for female interpreters

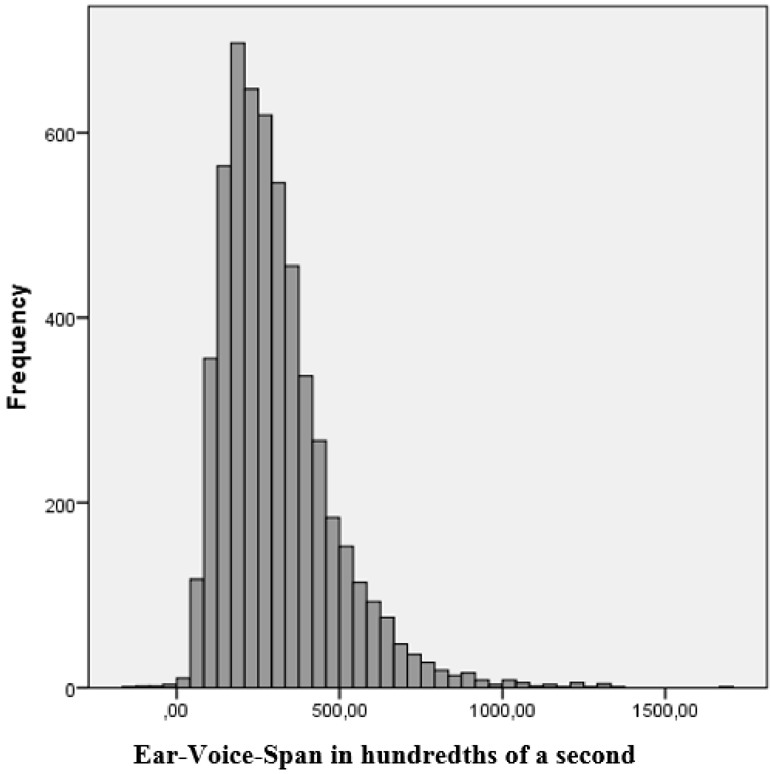


Figure 3. Histogram EVS for male interpreters

The Mann-Whitney U-test compares EVS in two groups: males and females. The Mann-Whitney test run on 10864 time tags indicated that the dependent variable, EVS, was not statistically different according to the sex variable ($U=14715834.500$, $p=0.410$). This finding does not reject the null hypothesis and the hypothesis of the first research question: there is no evidence in the corpus that women have a shorter EVS than men.

6.3 Influence of predictors on EVS

An exploratory multiple regression was first conducted on a nearly identical corpus of 180 interpretations where all predictors were forced simultaneously into the model. The data analysed are the average values per 10 seconds. Seven predictors with non-significant p-values for the β-coefficient were progressively removed: interpreter's length of segments between pauses (*p*=.924), original speaker's length of segments between pauses (*p*=.953), interpreter's speech/pause ratio (*p*=.115), original speaker's speech/pause ratio (*p*=.636), original speaker's average duration of silent pauses (*p*=.652), interpreter's average duration of pauses (*p*=.860), original speaker's delivery type (*p*=.153) and the interpreter's sex (*p*=.710), as already shown by the Mann-Whitney U test with a p-value of .819. In total, eight predictors were considered as not having a significant influence on EVS. Finally a hierarchical forced entry multiple regression was conducted with the eight remaining predictors and the results in table 3 show that they have significant p-values associated with the β-coefficient: interpreter's language, delivery rate, number of lengthenings, filled pauses and false starts, original speaker's language and delivery rate, as well as the original speaker/interpreter speaking time ratio.

Table 3. Influence of significant predictors on EVS

Predictors	β	*p*-value	Standardized β	R2	Effect size
Interpreter's number of lengthenings	15.42	<.001	0.09	0.101	0.11
Original speaker's language: English	-27.50	<.001	0.14	0.047	0.05
Original speaker's language: French	-47.37	<.001			
Original speaker's language: Dutch	0				
Interpreter's language: English	31.99	<.001	-0.10	0.032	0.03
Interpreter's language: French	35.02	<.001			
Interpreter's language: Dutch	0				
Interpreter's number of filled pauses	12.11	<.001	0.11	0.092	0.10
Interpreter's delivery rate	-3.34	<.001	-0.14	0.072	0.08
Original speaker's delivery rate	3.05	<.001	0.11	0.122	0.14
Original speaker's/interpreter's speaking time ratio	13.27	0.006	0.04	0.114	0.13

Interpreter’s number of false starts	9.04	0.014	0.04	0.093	0.10

Multicollinearity has not been detected in the final dataset (VIF values are under 5 and no Pearson’s correlation coefficient is above 0.5 among predictors) and the assumption for the regression are met (the assumption of independent errors is met with a Durbin-Watson of 1.271 and residuals are normally distributed). The analysis of standardized residuals show that outliers are not influencing the model (only 4 of data have absolute values above 2 and 1.8% above 2.5).

The EVS increases when the interpreter produces more lengthenings, filled pauses and false starts. The EVS decreases when the interpreter’s produces more words per 10s but increases when the source speaker produce more words per 10s. When the original speaker/interpreter ratio increases, the EVS also increases. Table 3 also shows that the scale of the standardized beta values are identical for the original speaker’s language and the interpreter’s delivery rate (respectively 0.14 and -0.14). The interpreter’s language, number of filled pauses and the source speaker’s delivery rate also have almost the same amplitude for standardized beta values (respectively -010, 0.11 and 0.11). Given that the languages are categorical variables, the beta values are presented differently. Dutch as a source language triggers the longest EVS. Compared to Dutch, French decreases the EVS by 47.37 centiseconds (0.47 seconds) and English decrease the EVS by 27.50 centiseconds (0.28 seconds). When it comes to the target language, Dutch triggers the shortest EVS. Compared to Dutch, French increases the EVS by 35.02 centiseconds (0.35 seconds) and English increases the EVS by 31.99 centiseconds (0.32 seconds). Effect sizes for the influence of predictors range from 0.03 to 0.14, with 0.02 being a small effect size and 0.15 a medium effect size (Cohen, 1988).

6.4 Language-specific sex differences in ear-voice span

Given the influence of the original speaker and the interpreter's language, we decided to focus on the EVS in each language pair in table 4.

Table 4. Sex difference and EVS values per language pair

Language	Sample	Mean	Median	SD	N	Mann-Whitney U test		
						U	*p*	Effect size
French into English	All	2.78	2.48	1.51	1632	260035.000	<.001	0.18
	Females	2.56	2.23	1.41	905			
	Males	3.05	2.81	1.58	727			
French into Dutch	All	2.69	2.40	1.46	1806	378890.500	0.019	0.06
	Females	2.72	2.46	1.32	827			
	Males	2.67	2.32	1.58	979			
English into French	All	3.16	2.81	1.66	1889	337859.000	<.001	0.21
	Females	3.49	3.21	1.69	915			
	Males	2.86	2.53	1.56	978			
English into Dutch	All	2.67	2.41	1.31	1973	426565.000	<.001	0.11
	Females	2.79	2.51	1.34	1020			
	Males	2.54	2.31	1.28	953			
Dutch into French	All	3.41	3.03	1.89	1803	380896.000	0.022	0.05
	Females	3.24	2.93	1.61	887			
	Males	3.57	3.13	2.12	916			
Dutch into English	All	3.46	3.13	1.78	1757	312563.000	<.001	0.16
	Females	3.25	2.79	1.85	866			
	Males	3.66	3.40	1.69	891			

Similarly to the previous Mann-Whitney U test, the data in table 4 come from the whole sample (N=108,164) and not from the average values per 10 seconds, since only the EVS is taken into account and not the other predictors. The median values for EVS vary from one language combination to another and go from shortest to longest: French into Dutch (2.40 seconds), English into Dutch (2.41 seconds), French into English (2.48 seconds), English into French (2.81 seconds), Dutch into French (3.03 seconds) and Dutch into English (3.13 seconds). Compared to the general median (3.03 seconds) only the combinations Dutch into English produces a longer EVS. The combinations English into French, French into English, French into Dutch and English into Dutch produce shorter EVS, and Dutch into French has the same median EVS as the general median EVS.

The Mann-Whitney U test was performed for each language pair (see table 4). The p-values were adjusted for Type I error using the Bonferroni method and the significance level is therefore equal to 0.008. The Mann-Whitney U tests gives the following results: the p-values show that EVS is significantly different between males and females for four languages pairs out of six, but gave no difference from French into Dutch and from Dutch into French. Male interpreters have a longer EVS in the English booth (effect size=0.18 from French and effect size=0.16 from Dutch), while female interpreters have a longer EVS from English (effect size=0.21 into French and effect size=0.11 into Dutch). However, the effect sizes mentioned above seem to indicate that these differences can be considered as negligible.

7. Discussion and conclusions

The results of the EVS analysis for the whole sample as well as for each language pair seem to indicate that there is no sex differences in EVS. While two language pairs show a longer EVS for male interpreters (Dutch into English with an effect size of 0.16, French into English with an effect size of 0.18), and two language pairs show a longer EVS for female interpreters (English into Dutch with an effect size of 0.11, English into French with an effect size of 0.21), the effect sizes are very low and these differences are therefore negligible.

Out of the sixteen predictors analysed in this study, eight have no influence on the EVS: interpreter's and original speaker's length of segments between pauses, interpreter's and original speaker's speech/pause ratio, original speaker's and interpreter's average duration of silent pauses, interpreter's sex and delivery type. It appears clearly that silent pauses (their length and frequency) have no influence on the EVS, contrary to Lee's findings (2002). The length of segments between pauses is also irrelevant and seems to indicate that the length of sentences in the original speech is not particularly relevant. Unlike Barik (1973), we did not find a significant influence of the type of delivery, which might be due to the fact that the information about delivery type was sometimes incomplete because of the lack of video data for some speeches, and to the fact that some speeches were partly read and partly impromptu (labelled as 'mixed').

The remaining eight predictors have a statistically significant influence on EVS, with small and medium effect sizes. While the effect size for original language is small (0.05), French seems to trigger the shortest EVS, while Dutch produces longer EVS, which is in line with findings indicating that SOV source languages require a longer EVS (Goldman-Eisler, 1972). The EVS increases when the number of uttered disfluencies (lengthenings, false starts and filled pauses) by the interpreter increases, which seems logical as disfluencies delay the output of the interpreter and therefore increase EVS. The effect sizes for lengthenings (0.11), false starts (0.10) and filled pauses (0.10) are medium. The EVS also increases when the original speaker's delivery rate increases and when the original speaker/interpreter speaking time ratio increases, which is in line with previous findings (De Groot, 1997; Gerver, 1969; Lee, 2002), showing that when the rate and quantity of input the interpreter needs to process increases, the cognitive effort increases and so does the EVS. The effect size for the original speaker's delivery rate (0.14) and the original speaker/interpreter speaking time ratio (0.13) are medium. When it comes to the last predictor, the interpreter's delivery rate, results show that it has a negative impact on the EVS and a small effect size (0.08). When the interpreter speaks faster, the EVS decreases, contrary to previous findings by Lee (2002).

Given that the ear-voice span is a good indicator of cognitive processes involved in SI and given the sex differences found in several cognitive tasks involved in SI, the aim of this research project was to determine whether sex differences in ear-voice span can be found for interpreters at the European Parliament in six language combinations. It was assumed that women would need to dedicate fewer cognitive resources to the interpreting task and would therefore have a shorter EVS than men. The results did not confirm the hypothesis as no clear sex difference was found. While p-values were significant for four language pairs, the effect sizes are negligible and the observed differences are not homogeneous, as two language pairs show a longer EVS for women while two present a longer EVS for men.

These results reflect the complex nature of the interpreting process and the difficulty to determine which factors influence the EVS. Indeed while this study revealed that several predictors have an impact on the EVS, their influence is often small and suggest that other predictors also play a role. For example the following predictors have been identified in previous literature and have not been analysed in this research project: the syntactic complexity and lexical density of the input and output, as well as the interpreter's personal preferences and strategies, level of preparation and experience. This also confirms the high variability of interpreting data. In 2003, Lamberger-Felber found that on a whole range of measures (errors, EVS, lexis, etc.), the interpreters varied greatly and their performance differences did not correspond to the differences in text types. In other words, there was much individual variability in interpreting which could not be attributed to obvious external factors. Indeed interpreters may behave very differently and employ a variety of different processes, strategies and norms depending on their individual strengths and weaknesses.

Addendum to Chapter 1: Predictors of Ear-Voice Span, a corpus-based study with special reference to sex

5. Materials and methods

5.4 Predictors

[...]

After the Mann-Whitney U test performed to identify potential sex differences, a generalized linear mixed model was conducted with all predictors in order to determine their individual significance. Mixed models allow for each interpretation to be included as a random variable in order to control for idiosyncratic effects. The skewness of EVS does not prevent a regression to be carried out as only the residuals need to be normally distributed. The results are presented through the β-coefficient, which indicates the individual contribution of each predictor to the regression model. In order to estimate the goodness of fit of the model, the marginal R^2 and the conditional R2 are also mentioned. The marginal R2 is the variance explained by the fixed effects, while the conditional R2 is the variance explained by both the fixed and the random effects.

6. Results

6.2 Sex differences in ear-voice span

[...]

The Mann-Whitney U-test compares EVS in two groups: males and females. The Mann-Whitney test run on 180 interpretations indicated that the dependent variable, EVS, was not statistically different according to the sex variable (U=4038.500, *p*=.974). This finding does not reject the null hypothesis and the hypothesis of the first research question: there is no evidence in the corpus that women have a shorter EVS than men.

6.3 Influence of predictors on EVS

A generalized linear mixed model was conducted with all predictors and the random effect “interpretation”. The data analysed are the average values per 10 seconds.

Table 3. Influence of significant predictors on EVS

Predictors	β	Std. Error	df	t-value	Sig.
Significant predictors					
Interpreter’s number of lengthenings	9.450	2.535	4094.248	3.728	<.001
Interpreter’s number of filled pauses	16.695	1.832	4145.983	9.113	<.001
Original speaker’s delivery rate	2.353	0.570	4150.142	4.127	<.001
Original speaker’s/interpreter’s speaking time ratio	20.320	9.872	4020.070	2.058	.040
Interpreter’s number of false starts	16.865	3.418	4061.297	4.934	.014
Original speaker’s language: Dutch	37.388	18.287	154.644	2.045	.043
Original speaker’s language: English	0				
Interpreter’s speech/pause ratio	1.946	0.527	4072.004	3.691	<.001
Original speaker’s length of segments between pauses	-5.183	1.293	4144.483	-4.007	<.001
Original speaker’s average duration of silent pauses	-1.688	0.770	4124.574	-2.193	.028
Interpreter’s average duration of silent pauses	14.118	3.669	4096.599	3.848	<.001
Not significant predictors					
Interpreter’s length of segments between pauses	-1.006	1.070	4117.387	-0.940	.347
Original speaker’s speech/pause ratio	0.793	0.882	4054.18	0.900	.368
Original speaker’s delivery type	65.971	80.669	151.153	0.818	.415
Interpreter’s sex	2.473	12.229	156.928	0.202	.840
Original speaker’s language: French	-15.286	17.938	155.970	-0.852	.395
Interpreter’s language; Dutch	-27.153	16.924	156.307	-1.605	.111
Interpreter’s language: French	5.067	17.087	156.794	0.297	.767
Interpreter’s delivery rate	-0.498	0.551	4141.591	-0.904	.366

The marginal R^2 is 0.113 and the conditional R^2 is 0.387. On top of the full model presented in Table 3, the influence of predictors was also measured in single models between each predictor and the EVS. Five predictors do not have significant p-values in the full model but did yield significant p-values in the single models: interpreter's length of segments between pauses with p=.347 (p>.001 for the single model), original speaker's speech/pause ratio with p=.368 (p=.029 for the single model), interpreter's language Dutch with p=.111 (p=.002 for the single model), original speaker's language French with p=.395 (p=.045 for the single model) and interpreter's delivery rate p=.366 (p<.001 for the single model). Three predictors do not have significant p-values in both models: interpreter's language French (p=.767), interpreter's delivery rate (p=.366), and as already shown by the Mann-Whitney U test, the interpreter's sex (p=.840).

Two predictors did not yield significant p-values when included individually in the model but are significant in the full model: interpreter's speech/pause ratio with p<.001 (p=.370 for the single model) and original speaker's length of segments between pauses with p<.001 (p=.344 for the single model). Seven predictors (and the original speaker's language to some extent) have significant p-values in both models: interpreter's number of lengthenings (p<.001), filled pauses (p<.001) and false starts (p=.014), original speaker's delivery rate (p<.001), original speaker's/interpreter's speaking time ratio (p=.040), original speaker's language Dutch (=.043), original speaker's average duration of silent pauses (p=.028) and interpreter's average duration of silent pauses (p<.001). Interaction effects were found only for two predictors out of sixteen (interpreter's length of segments between pauses with interpreter's average duration of silent pauses with p<.001). Given that these two predictors are significant in the model, they were not removed.

The β values show that the interpreter's number of lengthenings, filled pauses and false starts, the interpreter's speech/pause ratio and average duration of pauses, as well as the original speaker/interpreter ratio and original speaker's delivery rate are positively correlated with the EVS. However, the original speaker's length of segments between pauses and average duration of silent pauses are negatively associated with the EVS. Finally, Dutch as a source language triggers a longer EVS than English.

Language-specific sex differences in ear-voice span

Given that the corpus presents balanced datasets for each language pair and that some language pairs seem to have an influence on the EVS, we decided to focus on the EVS in each language pair in Table 4. The Mann-Whitney U test was performed for each language pair on the aggregated data per interpreter.

Table 4. Sex difference and EVS values per language pair.

Language	Sample	Mean	Median	SD	N	Mann-Whitney U test		
						U	p	Effect size
All language pairs	All	3.10	2.94	0.89	180			
	Females	3.10	2.97	0.90	90			
	Males	3.10	2.90	0.86	90			
French into English	All	2.87	2.78	0.82	30	79.000	0.174	0.25
	Females	2.65	2.46	0.82	15			
	Males	3.08	2.99	0.79	15			
French into Dutch	All	2.68	2.75	0.44	30	84.000	0.250	0.22
	Females	2.76	2.87	0.38	15			
	Males	2.59	2.55	0.49	15			
English into French	All	3.35	3.28	0.94	30	58.000	.024	0.41
	Females	3.72	3.67	1.07	15			
	Males	2.97	2.75	0.63	15			
English into Dutch	All	2.74	2.75	0.48	30	95.000	.486	0.13
	Females	2.83	2.73	0.54	15			
	Males	2.66	2.76	0.42	15			
Dutch into French	All	3.50	3.41	1.07	30	106.000	0.806	0.05
	Females	3.34	3.12	0.76	15			
	Males	3.65	3.52	1.31	15			
Dutch into English	All	3.45	3.26	1.01	30	48.000	.007	0.49
	Females	3.27	2.97	1.25	15			
	Males	3.63	3.45	0.70	15			

The *p*-values were adjusted for Type I error using the Bonferroni method and the significance level is therefore equal to 0.008. The Mann-Whitney U tests give the following results: the *p*-values show that EVS is not significantly different between males and females for five languages pairs out of six, but is significantly different from Dutch into

English, where male interpreters have a longer median EVS (3.45s compares 2.97s for women). The effect size is 0.49.

The median values for EVS vary from one language pair to another and two distinguishable groups seem to emerge. On one side, language pairs with EVS smaller than the global EVS (2.94s): English into Dutch and French into Dutch (both 2.75s) and French into English (2.78s). On the other side, language pairs with EVS above the global EVS: Dutch into English (3.26s), English into French (3.28s) and Dutch into French (3.41s).

7. Discussion and conclusions

The results of the Mann-Whitney U tests for the whole sample as well as for each language pair, and the results of the regression seem to indicate that there is no sex differences in EVS. While one language pair show a longer EVS for male interpreters (Dutch into English with an effect size of 0.49), the five remaining language pairs show no significant differences.

Out of the sixteen predictors analysed in this study, seven have no influence on the EVS: interpreter's length of segments between pauses, source speaker's speech/pause ratio, source speaker's delivery type, interpreter's language, interpreter's delivery rate and the interpreter's sex. Contrary to Lee's findings (2002), the length of segments between pauses produced by the interpreter and the interpreter's delivery rate seem irrelevant. Unlike Barik (1973), we did not find a significant influence of the type of delivery, which might be due to the fact that the information about delivery type is sometimes incomplete because of the lack of video data for some speeches, and to the fact that some speeches were partly read and partly impromptu (labelled as 'mixed').

The remaining predictors have a statistically significant influence on EVS. Dutch as a source language seems to trigger a longer EVS, which is in line with findings indicating that SOV source languages require a longer EVS (Goldman-Eisler, 1972). The EVS increases when the number of uttered disfluencies (lengthenings, false starts and filled pauses) by the interpreter increases, which seems logical as disfluencies delay the output of the interpreter and therefore increase EVS. The EVS also increases when the original speaker's delivery rate increases and when the original speaker/interpreter speaking time ratio increases, as well as when the original speaker's average duration of silent pauses decreases, which is in line with previous findings (De Groot, 1997; Gerver, 1969; Lee, 2002), showing that when the rate of input the interpreter needs to process increases, the cognitive effort increases and so does the EVS. When the interpreter's speech/pause ratio increases, the EVS is longer, showing that an increase in the quantity of output might also increase the cognitive effort. However, when the original speaker's length of segments between pauses increases, the EVS decreases. This could simply mean that if the segments between pauses are too short, the interpreter needs to wait and increase the EVS in order to receive enough input. The results also show that when the average duration of silent pauses produced by the interpreter increases, the EVS increases as well.

In this case, longer silent pauses can potentially be an indicator of an increased cognitive effort for the interpreter.

Given that the ear-voice span is a good indicator of cognitive processes involved in SI and given the sex differences found in several cognitive tasks involved in SI, the aim of this research project was to determine whether sex differences in ear-voice span can be found for interpreters at the European Parliament in six language combinations. It was assumed that women would need to dedicate fewer cognitive resources to the interpreting task and would therefore have a shorter EVS than men. The results did not confirm the hypothesis as no sex difference was found.

[...]

Chapter 2: Corpus study on the rendition of numbers in simultaneous interpreting, with special reference to sex

Abstract

Experimental studies show that numbers constitute a challenge for simultaneous interpreters mostly because of their low predictability and density of information. While meta-analyses suggest that the reported cognitive differences between the sexes are often exaggerated, a female advantage has been found in individual studies for tasks that are crucial to interpreting. Assuming that women therefore need to dedicate fewer cognitive resources to the interpreting task, this paper's hypothesis is that female interpreters will have more available resources to deal with the complex task of rendering numbers and will therefore make fewer errors than men. This article relates a rare corpus-based study on the rendition of numbers by male and female interpreters at the European Parliament against the background of other potential predictors. The data sample consists of 180 source texts and interpretations in six language pairs (both from and into French, English and Dutch). The results did not confirm the hypothesis, as sex does not appear to be a significant predictor of the rendition of numbers but confirm that omissions tend to happen more frequently when the Ear-Voice Span is longer.

Keywords: simultaneous interpreting; numbers; corpus; sex differences

1. Introduction

This paper analyses the possible effect of sex, as well as other predictors, on the rendition of numbers. Numbers constitute a particular challenge in simultaneous interpreting (Alessandrini 1990; Chmiel 2015) with error rates reaching 40% for professional interpreters (Korpal 2016; Korpal & Stachowiak 2018; Timarová 2012;). Numbers are characterised by low predictability and redundancy (Braun & Clarici 1996; Gile 1995; Mazza 2001; Mead 2015; Pinochi 2009; Seeber 2015) which makes them almost impossible to anticipate or paraphrase (Jones 2002; Pinochi 2009). Moreover, several variables such as high delivery rates (Korpal 2016) and differences in number syntax between source and target language (Pinochi 2009) seem to increase the error rates. Given this difficulty, professional interpreters apply different strategies to deal with numbers: shortening their EVS to keep as close as possible to the source number and not overload the memory (Jones 2002; Setton 1999; Timarová et al. 2014); writing down the number (Jones 2002; Mead 2015; Setton 1999); relying on external visual input (e.g. PowerPoint) or notes from the booth colleague (Desmet et al. forthcoming; Lamberger-Felber 2001; Mead 2015; Seeber 2012) and switching from intelligent hearing to literal hearing (Pinochi 2009; Seleskovitch 1975).

Numbers have mostly been studied in experimental designs. However, as Korpal & Stachowiak (2018) note, experiments are faced with an 'age-old problem' that some behaviours may be triggered by the laboratory situation. They therefore suggest examining how conference interpreters deal with numbers in a natural working environment, i.e. in one of the institutions of the European Union. With this paper, we intend to answer that call, drawing on a parallel acoustic aligned and time-tagged sub-corpus of the European Parliament Interpreting Corpus Ghent (EPICG) which includes transcriptions of speeches and their interpretations recorded during plenary sessions of the European Parliament. It is therefore one of the rare corpus-based studies on the rendition of numbers in simultaneous interpreting. There are some limitations attached to the analysis of numbers in real-life environment data. Firstly, speeches uttered during the plenary sessions of the European Parliament are generally short (1 to 6 minutes) and contain few complex figures, which drastically reduces the variety. Secondly, the conditions cannot be controlled or verified. In this case, this means that interpreters can write down the number or benefit from external help (for example from a booth colleague, the written speech or a PowerPoint presentation) without the researchers knowing. As a consequence, the error rate in a pilot corpus-based study is close to 18% for interpreters at the European Parliament (Collard & Defrancq 2017), i.e. less than half the error rate reported in most experiments.

The present data reflect the real working conditions of simultaneous interpreters and have the advantage of presenting a large diversity of interpreters, languages, source speakers and topics that allow for a representative study to be conducted. Moreover, the institutionalised setting of the debates and the mandatory accreditation tests for interpreters ensure that the data are fairly homogeneous in regard to the interpreter's working conditions and baseline quality. Moreover, even though the error rate is low, interpreters do make mistakes when rendering numbers in real-life environments.

This paper is part of a broader research project on cognitive sex1 differences in simultaneous interpreting (SI). While the existence of cognitive sex differences in general has not been proven and meta-analyses show that males and females are more similar than they are different (Hyde & Linn 1988), some specific verbal skills relevant to SI tend to show a female advantage (see Section 'Sex differences in cognitive tasks'). Only few researchers look at the influence of potential cognitive sex differences in simultaneous interpreting, but they all have reached promising results. In small preliminary corpus-based studies, Defrancq (2013) and Baes (2012) discover longer EVS for female interpreters at the European Parliament. In an experimental study on 11 professional interpreters, Cecot (2001) finds that women produce an average of 10.7% more filled pauses than men and men produce an average of 14.9% more silent pauses than women. Analysing recorded courtroom proceedings, Mason (2008) finds that female consecutive interpreters omit linguistic features that signal deference more than men, while male interpreters omit more politeness markers. While these findings are restricted in significance and cannot be generalised to all interpreters (as they focus on specific populations such as interpreters at the European institutions or in courts), they suggest a potential sex dimension to the interpreting process that deserves further exploration. Indeed, if sex appears to influence the way interpreters work, it would need to be more systematically included in interpreting research.

While studies on gender differences, i.e. differences in the interpreter's perception and role, are also rare (Baer & Massardier-Kenney 2016), results are inspiring. In corpus-based-studies, Magnifico & Defrancq (2016; 2017) have discovered that female interpreters at the European Parliament use more hedges and downtone fewer unmitigated face-threatening acts than male interpreters (with large effect sizes, respectively Φ=0.4 and 0.49). Analysing 704 answers from professional interpreters to a survey on the perception of quality, Pöchhacker & Zwischenberger (2010) show that

---

[1] For the purpose of this paper, differences between males and females will be described as sex differences. While the expression 'gender differences' is commonly used, it generally refers to an individual's self-conception and role within society. In fact, gender differences studies tend to focus on communicative and linguistic differences (Chambers & Trudgill 1998; Coates 1993). The present study, however, focuses on the cognitive aspects and therefore, as for most studies described here, only takes the subjects' biological sex into account and makes no assumptions on their gender.

female interpreters rate others' performances more generously and value a lively intonation in interpreting more than male interpreters. However, Angelelli (2004) found no gender difference in the analysis of interpreters' perceptions of roles and behaviours in practice. Studies on cognitive sex differences can complement studies on gender either by offering additional explanations to the differences found or by showing that gender differences are not based on cognitive differences but are mostly societal, cultural or educational. Studies on sex and gender differences can also help eliminate stereotypes. For example, they could contribute to explaining why few men opt for the interpreting career. Indeed, female students clearly outnumber male students in interpreting training (Hickey 2018; Lim 2005; Ryan 2015) and this imbalance remains unexplained. Miller and Halpern (2014) give a partial explanation by indicating that individuals choose their education and career partly based on their relative cognitive strengths. As a consequence, studies on sex differences in verbal tasks showing a female advantage might have an impact on men's decisions, as they might not opt for interpreting based on their potentially ill impression of a female predisposition for interpreting. More importantly, their performance could be negatively impacted by this perception, given that people's performances appear weaker when they are told that the other sex performs better at the task (Spencer et al. 1999).

Finally, given their difficulty and the strategies needed to overcome it, numbers play a key role in interpreting training. While it is unlikely for sex-specific courses to be organized, the existence of sex differences in interpreters' performances would suggest that individual skills play a significant role in the way interpreters work. Studying sex differences is an important step to understand how professionals find different ways to solve similar problems and can help design more effective training curricula. Indeed, Halpern et al. (2007) suggest that if men and women have different ways of solving problems, training can be adapted to correspond to their most efficient cognitive process and to allow more flexibility.

With that in mind, the aim of this study is multifold. Firstly, it is to observe the rendition of numbers in a real environment, acknowledging that interpreters might have received some kind of help. Secondly, it is to analyse potentially diverging behaviours between male and female interpreters dealing with numbers, assuming that the level of help both sexes receive is equivalent. This study will focus on the categories of rendition and on the ear-voice span of interpreters when dealing with numbers, as it is one of the strategies applied. Finally, we will take into account the influence of other predictors on the rendition of numbers: the interpreter's sex, source delivery rate, source speaker's number of silent pauses per minute, source speaker's average duration of silent pauses per minute, number's EVS, mean EVS, source speech type, nature of the number, complexity of the number, source and target languages.

In the section 'Numbers in interpreting', the complexity of rendering numbers in interpreting, as well as the strategies developed by professional interpreters are discussed. The Section ‘Sex differences in cognitive tasks’ covers the literature on sex differences for several cognitive skills relevant to simultaneous interpreting. Section 2 ‘Methodology’ presents the corpus and the methodology used to identify and measure the numbers, their rendition and the potential predictors. The results of the statistical analyses are reported in Section 3. Finally, Sections 4 and 5 presents the conclusion and discussion of these results.

1.1 Numbers in interpreting

Simultaneous interpreting is a cognitively demanding task involving several processes: speech comprehension and production, memory, attention/resource allocation and coordination (Pöchhacker 2015). After conducting neuroimaging studies during simultaneous interpreting, Hervais-Adelman et al. (2015: 1) confirm that simultaneous interpretation 'places extreme demands on the cognitive control of language and on verbal working memory and attention'. The constraint of simultaneity and the limited capacity make interpreting particularly challenging. The most popular model put forward to chart cognitive efforts during interpreting, and simultaneous interpreting is Gile's (1995; 2018). In his model, tasks are divided into four efforts, which compete for limited attentional resources (also called processing capacity): the listening and analysis effort (the detection and identification of stimuli and the assignment of a meaning to what is heard), the short-term memory effort (the storage mechanism where information is temporarily kept), the production effort (the planning, production and self-monitoring of the speech in the target language) and the coordination effort (the management of attentional resources to the three other efforts, given that these resources are believed to be limited). Gile (1995; 2018) also suggests that omissions or errors are caused not only by specific difficulties in the source speech but because attentional resources required to perform the task adequately were not available for a particular comprehension, memory or production task at a time when they were needed. The reason behind this lack of availability is that simultaneous interpreters tend to work close to cognitive saturation (referred to as the tightrope hypothesis).

There is a broad consensus among interpreters that numbers constitute a challenge in interpreting and are cognitively demanding (Alessandrini 1990; Chmiel 2015). 65% of professional interpreters participating in a survey on stress at work mentioned numbers as a source of stress (Alessandrini 1990). Moreover experimental studies on the rendition of numbers by interpreters show that error rates (including omissions) reach 40% for professional interpreters (Korpal 2016; Korpal & Stachowiak 2018; Timarová 2012) and 40 to 70% for trainees (Braun & Clarici 1996; Korpal 2016; Mazza 2001; Pinochi 2009). Scholars have identified several reasons why interpreting numbers is challenging. Firstly, numbers lack a conceptual representation (Seeber 2015; Timarová 2012) and are characterised by low predictability as the quantity expressed can often only be understood the moment it is uttered by the speaker (Braun & Clarici 1996; Mazza 2001; Mead 2015; Pinochi 2009). Therefore, they are almost impossible to anticipate (with several exceptions, such as well-known dates). Secondly, numbers are highly informative (Alessandrini 1990) and lack redundancy (Gile 1995; Seeber 2015) as every component of a number is a meaningful unit representing only one particular meaning. This often prevents interpreters from using strategies such as paraphrasing or reformulation (Jones 2002; Pinochi 2009), except when numbers are part of a metaphor, or are culturally 'lucky' or ‘magic’ (Chesterman 1997; 10). Moreover, several variables can complicate the rendition of numbers. Korpal (2016) finds that high delivery rates trigger more errors (43%) than slow delivery rates (30%) for experienced interpreters. Pinochi (2009) suggests that differences in number syntax between source and target language can exacerbate the error rate, as interpreters have to swap the order of some of the units (e.g. between English and Dutch).

Braun and Clarici (1996) put forward a rather comprehensive list of errors, which was replicated by several researchers, sometimes in simpler versions (Desmet et al. forthcoming; Mazza 2001). The most common type of rendition (besides complete rendition) is the complete omission (Desmet et al. forthcoming; Mazza 2001). Given the complexity of dealing with numbers, omissions could also be seen as a strategy adopted by interpreters to catch up when they are lagging behind or when they consider that the cognitive resource dedicated to rendering the number might be detrimental to the rest of the incoming speech. The second most frequent category of rendition is approximations, i.e. an erroneous rendition when the number is of the right order of magnitude and not far away from the original, sometimes accompanied with phrasal elements, such as ‘about’ (Desmet et al. forthcoming; Mazza 2001). This type of rendition can also be seen as a conscious decision to minimise any risk of compounding the error. The interpreters are aware that they are missing some specific information, but since they cannot store all of it, they choose to simplify the number. Braun and Clarici’s typology also includes

- lexical errors, i.e. renditions in which specific components of the number are wrong, while the order of magnitude is correct;
- syntactical errors, i.e. renditions in which the specific components are correct, while the order of magnitude is wrong;
- inversions, i.e. renditions in which specific components are permutated;
- errors of phonemic perception, i.e. renditions where the error is related to a phonemically wrong perception of the number;
- self-corrections, either positive or negative.

It should be noted that Braun and Clarici’s categorisation is not systematic, as it includes purely formal criteria, but also a causal one (phonemic perception). Overlaps are therefore not excluded: one single case can be an example of both lexical error and phonemic perception error.

Considering that the study of number renditions in interpreting should not only focus on the numbers themselves, Ooms (2008) complements Braun and Clarici’s typology with the category ‘context’, referring to cases where the number is rendered correctly but combined with a wrong unit. Indeed when numbers are interpreted simultaneously, a large amount of cognitive capacity is spent on the processing of the numbers themselves, leaving less capacity for neighbouring items, such as the element a number refers to

(Jones 2002). Consequently, interpreters might tend to resort to approximations or omissions when it comes to those number-accompanying items (Jones 2002), which might be rendered with less accuracy. This loss of accuracy, or inefficient cognitive capacity management in the event of a difficult stimulus, is often referred to as the spillover effect (Gile 1995/2009; Seeber 2011; Shlesinger 2000). The term has been first used in psychology and psycholinguistics to describe a task switching effect in which a secondary task is delayed as a result of a cognitively taxing primary task (Meiran et al. 2002; Zillmann 1971).

Previous experiments also show a relationship between the type of error and the features of the source number, as most errors seem to occur with large numbers and decimals (Desmet et al. forthcoming). Most authors therefore make a distinction between small and large numbers and between dates, decimals and other numbers (Mazza 2001; Pinochi 2009).

Strategies

Several experimental studies compare the success rate of the rendition of numbers between professional and student interpreters and found that professional interpreters have better total accuracy scores than trainees (Korpal & Stachowiak 2018). Timarová et al. (2014) suggest that professional interpreters perform better than students because professionals keep a shorter ear-voice span (i.e. the lag time between a word uttered by the source speaker and the corresponding word uttered by the interpreter) when dealing with numbers, in order to stick as close as possible to the source speaker and reduce the amount of time during which the number must be kept in memory, i.e. the phonological store (Baddeley and Hitch 1974). Jones (2002) also suggests that when faced with numbers with two or more components, interpreters keep as close as possible to the speaker and Setton (1999) observes that errors typically occur when the EVS is more than 3–4 seconds. The strategy of reducing EVS seems logical as Gile (2008) states that a short EVS potentially means that the interpreter is saving working memory capacity.

Note-taking is one of the universally recommended strategies (Mead 2015; Setton 1999). Interpreters are advised to stop delivery of the target text as soon as they hear a number, write down that number and read it off while starting up delivery again. Mazza's study (2001) seems to prove that interpreting of numbers is more accurate when interpreters jot down the number. However, Mazza (2001) notes that taking notes when dealing with large numbers (i.e. with four or more digits) might also have a detrimental effect on their rendition, probably because large whole numbers are often too dense even to be successfully noted. In other words, any advantage in terms of reduced demand on working memory is offset by the extra processing capacity required for note-taking. Therefore, assistance by the booth colleague in writing down numbers and visual input provided by the speakers (such as a copy of the speech to be used in the booth, or the projected presentation slides) can also be beneficial (Mead 2015). Seeber (2012) conducted a study in which interpreters were presented with auditory stimuli including numbers and a pre-recorded video of the speaker's slides. Small numbers were presented by the speaker, who counted on their fingers, and large numbers were on the slides. He observed that interpreters did rely on the visual input when interpreting the numbers. Lamberger-Felber (2001) reports a significant increase in number and name accuracy (53 to 68% fewer errors) in an experiment where interpreters are provided the text of the speech in the booth, compared to when they do not have the text at their disposal. Desmet et al. (forthcoming) show that simulated technological support projecting the articulated number on a screen improves overall accuracy on numbers from 56.5 to 86.5% in a student population, reducing the number of errors by two thirds. They also show that technological help is most helpful in reducing errors on complex numbers and decimals, the two categories that are most often interpreted incorrectly, and that the occurrence of approximations drops by almost 90 percent when support is available.

Finally, changing the listening strategy seems to be another effective coping strategy. Following Seleskovitch (1975), Pinochi (2009) advocates a switch from intelligent hearing (i.e. taking into account the context to draw inferences) to literal hearing (i.e. paying attention to the item in isolation).

As the difference between professional and student interpreters in the rendition of numbers suggests, these strategies are not innate but acquired with practice and experience. Handbooks on interpreting training do generally include specific sections on the rendition of numbers (Gillies 2013; Setton & Dawrant 2016) and a specific website was created by a conference interpreter, Anton Klevansky, for the training of numbers in interpreting (http://www.numerizer.pro). Therefore it seems crucial that the rendition of numbers and the ad hoc strategies be taught as a skill during interpreting training, in order to help students develop the necessary strategies, and most specifically learn to process multimodal stimuli (auditory vs. visual) when receiving visual assistance (Korpal & Stachowiak 2018).

1.2 Sex differences in cognitive tasks

Several individual studies suggest that women have better verbal abilities, but these claims are often exaggerated and meta-analyses tend to show that males and females are much more similar than they are different (Hyde 2005; Hyde & Linn 1988; Miller & Halpern 2014). While meta-analyses fail to determine whether sex differences exist in general cognitive abilities, Hyde and Linn (1988) do suggest more fine-grained approaches are needed targeting specific abilities 'mov[ing] away from the old model of intellect that specified only three rather general cognitive abilities – verbal ability, mathematical ability, and spatial ability' (p 33). Similarly, this research project aims at exploring the influence of one specific predictor, i.e. sex, on one specific task, i.e. the simultaneous interpretation of numbers at the European Parliament. The goal is therefore not to determine whether cognitive sex differences exist in general or for populations other than simultaneous interpreters. Since simultaneous interpreting is the focus of this study, the literature review will only cover potential sex differences in skills involved in this activity, contrary to meta-analysis, where various and diverse tasks are included (e.g. spelling, reading, writing and vocabulary).

When it comes to verbal abilities in general, Hyde and Linn (1988) suggest that speech production is the only potentially significant sex difference, with an effect size of 0.33 and a higher score for women. However, several statistical sources would consider this effect size to be small (Ferguson 2009; Mellinger & Hanson 2017). Moreover, other verbal tasks show no significant sex difference (vocabulary with an effect size of 0.02 and anagrams with an effect size of 0.22) and/or give higher values for men (analogies with an effect size of -0.16). Several studies have concluded that women have greater verbal fluency, i.e. the ability to retrieve specific information within restricted search and time parameters, for example the ability to generate words beginning with a single letter in one minute (Herlitz et al. 1997; Loonstra et al. 2001; Maitland et al. 2004). Verbal fluency is believed to play a key role in interpreting (Stavrakaki et al. 2012). A female advantage in generating synonyms has also been found (Hines 1990). Women also seem to have an advantage when it comes to identifying spoken words. Indeed, Aerts (2013) found that women display a larger sensitivity to the phonemic contrasts during auditory phoneme discrimination and showed more differentiation in real word-pseudoword dissociation, while Keith et al. (2008) have found that women understand and react faster to the information they receive (i.e. processing speed). Similarly, a female advantage has been found both in pre-lexical and lexical processing (Majeres 1999), as well as for perceptual speed (the ability to compare or recognize items) (Born et al. 1987; Hedges & Nowell 1995). When it comes to verbal memory capacities, several studies have found that females perform better at episodic and some aspects of semantic memory tasks (Herlitz et al. 1999; Kramer et al. 1997; Maitland et al. 2004). A female advantage in immediate and delayed free recall and on recognition tasks with verbal and visual components has also been found (Kimura & Seal 2003; Trahan & Quintana 1990). However, Harness et al. (2008) report higher scores for males in a study on recall combined with a distraction task carried out on students. Other studies report better scores for women in the Rey Auditory Verbal Learning Test (free recall of two lists of nouns) and the Verbal Paired Associates test (immediate and delayed recall of word pairs) (Bolla-Wilson & Bleecker 1986; Gale et al. 2007). Unfortunately, most of these studies do not mention effect sizes (or do not give sufficient information for them to be calculated by the reader) and when effect sizes are mentioned, they are generally small, ranging from 0.3 to 0.6. While some of these results might therefore be exaggerated, we can suggest that female interpreters have several advantages when it comes to the listening/analysis, short memory and production efforts in interpreting.

Of particular importance to this study is the skill to recall numerical information. Early studies showed that women perform significantly better than men on digit-span tasks, i.e. tasks for which participants are asked to recall series of digits (Jensen & Reynolds 1983; Kail & Siegel 1978). However, more recent research does not confirm female superiority in the sense that women usually score better on the tasks, but differences remain at non-significance levels (Bermingham et al. 2013; Lynn & Irwing 2008; Solianik et al. 2016). It should be noted that although digit-span tasks are highly standardized and are, therefore, particularly well-suited for replication, they test a very specific skill which is unrepresentative for most human activities, including simultaneous interpreting. In simultaneous interpreting, numbers rarely come in a bare list. They are embedded in text, even in experimental research. Therefore, results from digit-span tasks cannot simply be extrapolated to complex and realistic settings such as simultaneous interpreting in the context of the European Parliament. Rather, our findings could shed light on the discussion from the point of view of situated performances, i.e. the performance of a particular task, such as number recall, embedded in a context which is representative of a professional activity.

1.3 Research question and hypotheses

Simultaneous interpreting is a complex task where several processes compete for limited cognitive resources. Numbers are considered cognitively demanding for interpreters as they prevent them from using common strategies (e.g. anticipation) which preserve working memory capacity. Sex differences have been observed for several of the processes involved SI, and more specifically short-term memory tasks, and most studies show a female advantage. Therefore, our main research question is: do these sex differences in cognitive skills influence the rendition of numbers? If we assume that women perform better than men on tasks linked to the processes involved SI, and more specifically short-term memory tasks, they have more cognitive resources available when dealing with the rendition of numbers and will therefore render more numbers than men. Indeed omissions and errors are believed to be caused by a lack of attentional resources required to perform the task (Gile 1995; 2018). Secondly, if men have fewer cognitive resources and memory capacity available when dealing with numbers, we can assume that they will resort to one of the strategies applied by professionals more often, i.e. present a shorter EVS.

In all, eleven predictors are included in the analysis: the interpreter's sex, source delivery rate, source speaker's number of silent pauses per minute, source speaker's average duration of silent pauses per minute, number's EVS, mean text EVS, source speech type, nature of the number, complexity of the number, source and target languages.

2. Methodology

2.1 Corpus-based interpreting studies

The vast majority of studies on numbers in interpreting is conducted in the framework of an experimental research design, which offers the advantage of controlled conditions. On the other hand, corpus data are produced by professionals in a real-life environment and reflect the interpreting activity in a way experimental data cannot. Shlesinger (1998) therefore calls for more corpus-based studies in interpreting in order to reinforce the empirical foundations of interpreting research. When corpora are available online, results can be reproduced and studies replicated.

Thanks to new technologies and tools such as EXMARaLDA (Schmidt & Wörner 2009), Praat (Boersma 2001) and SpeechIndexer (Szakos & Glavitsch 2007), the time-consuming compilation, transcription and analysis of interpreting data is simplified. Moreover, the plenary sittings and some of the committee sittings of the European Parliament can now be more easily retrieved from the website of the European Parliament, with access to important information about the speaker (political group and function, age, sex) and the speech (topic, time of the day, source speech type). The inclusion of this information about the context of the speech is essential in order to make sure that samples are not entirely taken out of the context in which they occurred (Diriker, 2004; Duflou, 2016). Several universities have compiled their own corpora, starting with the European Parliament Interpreting Corpus (EPIC), compiled at the University of Bologna (Bendazzoli & Sandrelli 2005), CoSi for consecutive and simultaneous interpreting (House et al. 2012) and DiK for dialogue interpreting in public service settings (Bührig et al. 2012) at the University of Hamburg, EPICG (European Parliament Interpreting Corpus Ghent) at Ghent University (Bernardini et al. 2018), and others at the universities of Rome, Trieste, Posnan, Louvain-la-Neuve and Saarbrücken.

2.1.1 The corpus used

The European Parliament Interpreting Corpus Ghent is a parallel corpus that consists of speeches and their interpretations recorded during plenary sessions of the European Parliament. Source and target texts are acoustically aligned on the basis of pauses with the transcription tool EXMARaLDA Partitur-Editor. For the transcriptions, the Valibel instructions (Bachy et al. 2007) were adapted to facilitate the machine readability of the texts. The detailed compilation, transcription and annotation process is described in Bernardini et al. (2018).

For this study, 30 source speeches of each of the six following combinations were selected: English-French, French-English, English-Dutch, Dutch-English, Dutch-French and French-Dutch. The inclusion of Dutch, for which number order is inverted compared to French and English (twenty-one becomes een-en-twintig) allows for the analysis of the influence of numerical syntax on the interpreter's rendition. The corpus comprises more than 14 hours of interpreted speech and a total word count of 108,245 interpreted words.

We did not have access to the interpreters' identities and were therefore faced with two methodological challenges for the selection of the speeches and their interpretations. Firstly, sex had to be determined on the basis of the properties of the recorded voices only. Human listeners are able to identify speaker sex with accuracy of over 95%, even on the basis of vowels pronounced in isolation. Human identification of sex is therefore considered a reliable method (Lass & Puffenberg 1971). For this study, our goal was to reach a set of 15 male and 15 female interpretations for each language pair2. To this end, we first determined sex using the manual sample of speeches on the European Parliament's website. Our classification was then verified by the transcriber and at least two additional reviewers. This process yielded an inter-rater agreement of 99.4%, with the three assessors diverging on only one interpretation. It was concluded that the disagreement came from a human encoding mistake. The three assessors finally agreed on all interpretations. In order to complement the human process, the identification of sex was complemented with a speaker diarization software (LIUM_SpkDiarization, Rouvier et al. 2013), which agreed on all identified sexes except for 8, reaching a human-machine agreement of 95.6%. However, two human assessors disagreed with the 8 sexes identified by the software and agreed with the human identification and it was decided to keep these data and trust the human identification. The software being optimised for different types of data (radio and TV shows), this divergence is not entirely surprising.

---

[2] According to the estimations of the Organisation Intersex International Europe, one percent of interpreters in the corpus could be intersex. Unfortunately, given the anonymity of data, there is no possibility of knowing whether interpreters are intersex. While we are aware that there is a chance that intersex interpreters are included in our samples, we consider that the large size of the sample means that this low probability is negligible.

Moreover, the corpus's audio data are complex (several speakers take the floor simultaneously and the quality is not always optimal), especially for the 8 sexes for which humans and software disagreed.

The second methodological challenge consisted of reaching a sufficiently varied sample of interpreters in order to ensure representativeness. Several methods were applied to reduce the risk of including multiple interpretations by the same interpreter in the corpus. The initial sampling of speeches from the European Parliament's website was done manually with the specific aim of limiting the number of interpreters with similar voices. Therefore, the period from which the recordings were taken stretches over 6 years and sessions were picked randomly to collect the interpretations. In the final dataset, 93 interpretations are drawn from 21 different dates in 2008, 47 interpretations from 17 different dates in 2009, 14 from 11 different dates in 2010, 9 from 9 different dates in 2011 and 16 from 4 different dates in 2013 and 1 in 2014. Moreover, the languages chosen for this study (English, French and Dutch) are sufficiently common to guarantee that they are covered by a large number of different interpreters.

We then made a first attempt at identifying similar interpreters in the corpus by distinguishing similar voices, but the high number of different interpreters made this task almost impossible. Therefore the software LIUM_SpkDiarization was used for the identification of identical interpreters in the corpus and the results are presented in Table 1. These results were considered reliable by two human assessors, but no specific human-machine agreement could be determined.

Table 1. Identification of identical interpreters by LIUM_SpkDiarization

Sex	Language	Number of interpreters identified		Total number of unique interpreters
		Twice	Three times	
Females	French	3	0	27
	English	3	1	25
	Dutch	4	0	26
Males	French	3	0	27
	English	2	2	24
	Dutch	2	0	28

Table 1 shows that the maximum number of interpretations by one and the same interpreter is 3, which is 10% of the interpretations included for a particular booth. While having identical interpreters in the data set could theoretically violate the assumption of independence of observations within each group (males and females), independence is unlikely to be compromised in our case. Indeed, given the large size of the corpus and the small number of potentially identical interpreters, the dependence of observations would only concern a small number of interpretations and the final corpus can be considered as representing a sufficiently diverse set of interpreters. Moreover, each interpretation is unique and performed in different conditions, even if performed by the same interpreter. Finally, data are not aggregated at the individual level, but at the group level (males and females) and the same interpreter cannot be included in both the male and the female group.

2.2 Identification of numbers and their rendition

Based on the literature and the fact that numbers in our data are rarely complex, a simple classification was adopted based on two criteria: the number of components and formal features of numbers. Two categories were distinguished for the first criterion: numbers with two or fewer components (e.g. one, seventeen, three thousand) and numbers with three or more components (e.g. two hundred fifty, two thousand and eight). As Vander Beken (2013) suggests, counting uttered components instead of digits or quantity expressed makes more sense for oral data. For example, while 'thousand' and 'two' are different in size, they only need the pronunciation of one component, and therefore require the same memory effort. It is important to note that the same number uttered in two different languages can belong to two different categories: e.g. 'ninety-nine' has two components and is in the category two or fewer, while its translation in French, 'quatre-vingt-dix-neuf', has four components and is in the category three or more. Given that this research project focuses on potential errors in the rendition of numbers, only numbers that have to be translated by other numbers are taken into account. For example, the 'second' in 'my second topic' is not included as interpreters might decide to render the expression by 'my next topic' without making a mistake.

For the second criterion, i.e. the nature of the number, three categories were identified in the literature and replicated here: dates (which include days, months and years), decimals and others (all figures that do not belong to the former categories). The decision to add information on the type of number is justified by two reasons. Firstly, unlike most numbers, dates are usually not deprived of context and have a higher predictability than other numbers, as they can often be anticipated thanks to the context or prior knowledge (Vander Beken 2013), e.g. the Agenda 2020. Secondly, decimals are considered more troublesome than other types of numbers (Desmet et al. forthcoming; Mazza 2001) and therefore deserve their own category.

As mentioned in the literature, several classifications have been put forward by researchers for the rendition of numbers. One of the advantages of corpus data is the large diversity of interpreters, which allows for different interpreting styles and strategies to be included. Therefore, the categorisation adopted for this paper attempts at representing the large diversity of different renditions in the clearest way possible. Unfortunately, the previous classifications are not always systematic and contain some overlaps and some adaptations were needed for this study. The present categorisation was created with the following goals in mind: cover all the possible types of rendition in a systematic way, in order to ensure replicability and avoid overlaps. Moreover, several previous categories could actually be considered as subcategories rather than entirely new categories and were therefore included as such in the present classification. Finally, previous authors often include omissions as a type of error. Omissions being often associated with strategies in simultaneous interpreting, they are taken here as a specific category and not as errors. Moreover, a distinction is made between implicit and explicit approximations, as explicit approximations suggest that the interpreters are conscious of their error. Given the small sample size for numbers, the different types of erroneous renditions were compiled into three main categories in order to allow for representative analyses to be carried out: unrelated substitutions, related substitutions and approximations. We can argue that there is an increasing level of accuracy with these three categories: approximations are closest to the original number, while unrelated substitutions are the least faithful category. The types of renditions were manually added to the transcription. The final categorisation with explanations and examples is presented in Table 2.

Table 2. Categories of rendition

Category	Sub-category	Explanation	Example
Complete rendition		The number is correctly and entirely rendered	1999 => 1999
	Interpreter’s correction	The interpreter first made a mistake but then gave the correct rendition	
Complete omission		The number is completely omitted or is replaced by a word or sentence	1000 => a lot
Approximation		The order of magnitude is correct, but the number has been rounded up or down	1864 => 1800
	Explicit approximation	The interpreter specifies that the number is an approximation	125 => more than 100
Related substitution		The number has been replaced by another number which has some resemblance to the original number	See subcategories
	Phonological substitution	The rendered number has a phonological resemblance with the original number	14 (fourteen) => 40 (forty)
	Syntactic substitution	The rendered number has a syntactic resemblance with the original number	47 => 470

	Partial inversion	The main components have been rendered but some have been inverted	164 => 146
	Complete inversion	All components have been permuted	64 => 46
	Contextual substitution	The number has been correctly rendered, but not its unit	100 percent => 100 dollars
Unrelated substitution		The whole number or some parts of the number have been substituted by a number with no resemblance	See subcategories
	Partial substitution	One part of the number has been substituted	72 => 73
	Complete substitution	The whole number has been substituted	58 => 140
Interpreter's repetition		The interpreter repeats the same number, while it was not repeated in the original	
Interpreter's addition		There was no figure in the original speech, but the interpreter still uttered a number	Few => two
Interpreter's wrong correction		The interpreter made a mistake, tried to correct him(her)self but made a second mistake.	

The distinction between unrelated partial substitutions and related substitutions might not always be clear-cut, but the key is that there must always be some degree of resemblance with the source number for related substitutions, while this is not the case for unrelated substitutions. It is worth noting that the category 'contextual substitution' can occur simultaneously with other types of erroneous renditions. Moreover, for some categories, and automatically their subcategories (i.e. approximation, related and unrelated substitution, except for contextual substitution), the proportion of errors compared to the total number of elements is added. For example, if twenty-two is rendered by twenty-one, it is an unrelated partial substitution with a 50% of error (one erroneous component out of two). The categories 'interpreter's repetition' and 'interpreter's addition' do not automatically constitute errors and are therefore not included in the analyses. Moreover, the category 'interpreter's correction' is transformed into 'complete rendition' for the analyses.

2.3 Identification of predictors

In total, eleven predictors were included in the analysis: the interpreter's sex, source speaker's delivery rate, source speaker's number of silent pauses per minute, source speaker's average duration of silent pauses per minute, number EVS, mean EVS, source speech type, nature of the number, complexity of the number, source and target languages. Other predictors were mentioned in the introduction but could not be included in this analysis because they are not available. Indeed since the interpreters are anonymous, we have no information about their preferences or experience. Some of the predictors are directly derived from the literature; others will be explored for the first time in this study.

In the literature, variables identified as potentially having an impact on the rendition of numbers include source text delivery rate (Korpal 2016) and the language pair (Pinochi 2009). The source delivery rate is measured as the total number of words divided by duration of the speech, normalised by the minute. To complement data about the delivery rate, two variables about the pause patterns of the source speaker are added to the analysis. Firstly, the source speaker's number of silent pauses per minute which is the total number of silent pauses (i.e. an unfilled pause of more than 0.2s) for the whole speech normalised per min. Secondly, the source speaker's average duration of silent pauses per minute, which is the total duration of all silent pauses normalised per minute. The number type (date, decimal and others) and complexity (two or fewer components and three or more components) are also included as predictors, as the literature suggests that more errors are committed with decimals and large numbers (Desmet et al. forthcoming; Mazza 2001). Moreover, having a short EVS is considered as a strategy to reduce the risk of errors when dealing with numbers. Therefore two measurements of EVS were analysed: the EVS measured for each number and the mean EVS for the whole speech. The latter was measured by tagging source and target equivalent items every 10 word of the interpreter's rendition. The number EVS is measured between a number uttered by the speaker and the corresponding number in the interpretation. In both cases, the Ear-Voice span is measured by manually tagging the source item and its equivalent in the target language and measuring the time elapsed between the two in centiseconds. Two categories of rendition do not have an associated number EVS: interpreters' omissions (since there is no target item) and interpreters' additions (since the target number was not triggered by a source item). Another strategy mentioned when dealing with numbers is the use of visual input. Unfortunately, it is impossible for the authors to know whether the interpreter writes down the number or receives help from a booth colleague. However, thanks to the video attached to the source speech at the European Parliament, it is possible to see whether the source speaker improvises a speech or reads from a written document. When the source speaker improvises, the speech is categorised as ‘impromptu’ and it is unlikely for the interpreter to have received the speech in advance. However when the source speaker reads off a document, there is a higher probability that the interpreter has received the written speech and has numbers written down in front of her/him. As a consequence, the type of source speech will also be analysed as being impromptu, mixed (when the speaker alternatively reads off a document and speaks without looking at a document) or read, conforming to the encoding in Bernardini et al. (2008). However, some video data were not available and the information about source speech type is sometimes incomplete. Some predictors are manually added to each transcription (language, sex, type of source speech, nature of the number and complexity of the number) while others are automatically measured by a tailor-made script after having been manually identified in the text (both measurements of ear-voice span, delivery rate and pausing pattern).

A Chi-square test is performed on five types of rendition only in order to analyse the influence of sex on the rendition of numbers. The five categories are complete renditions, omissions, approximations, related and unrelated substitutions. Given the low occurrence of errors in the data set, conducting the analyses on all the categories and subcategories of rendition would not yield reliable results. Finally, the sex difference in the length of EVS for numbers was tested with a Mann-Whitney U test given that the distribution of EVS is not normal, but rightly skewed. For the influence of predictors on the rendition of numbers, logistic mixed models were performed on three categories of rendition: complete rendition, complete omission and erroneous rendition (the latter includes all other types of rendition). Logistic mixed models allow for each interpretation to be added as a random variable in order to control for idiosyncratic effects. The results are presented through the β-coefficient, which indicates the individual contribution of each predictor to the regression model. In order to estimate the goodness of fit of the model, the marginal R^2 and the conditional R^2 are also mentioned. The marginal R2 is the variance explained by the fixed effects, while the conditional R2 is the variance explained by both the fixed and the random effects.

When it comes to the analysis of the number EVS, it is important to mention that the data for complete omissions are not entirely adequate: the EVS is indicated as 0 for complete omissions while there is no number EVS measurement for this type of rendition. As a consequence, the analysis of the number EVS is also carried out without the category ‘complete omission’.

3. Results

3.1 Descriptive statistics

The corpus comprises 143 interpretations for which numbers were included in the source speech, source speakers produce 718 numbers in total, while the interpretations contain 729 numbers (11 numbers were either added or repeated by the interpreter). Numbers are divided into 6 categories: regular numbers, dates and decimals each with two or fewer and three or more components. Their distribution for all speeches, as well as for impromptu and read speeches only, is presented in Table 3.

Table 3. Categories of numbers

Speech type	Complexity	Numbers		Dates		Decimals		Total	
		#	%	#	%	#	%	#	%
All speeches	two or less	411	56%	65	9%	24	3%	499	69%
	three or more	81	11%	140	19%	8	1%	229	31%
	Subtotal	492	67%	205	28%	32	4%	729	100%
Impromptu speeches	two or less	61	55%	3	3%	6	5%	70	63%
	three or more	20	18%	21	19%	0	0%	41	37%
	Subtotal	81	73%	24	22%	6	5%	111	100%
Read speeches	two or less	259	55.5%	47	10%	13	3%	319	68.5%
	three or more	46	10%	97	21%	2	0.5%	145	31.5%
	Subtotal	305	65.5%	144	31%	15	3.5%	464	100%

For all speeches, the most frequent category is ‘regular’ numbers which amount to 67% of numbers, followed by dates (28%) and decimals (4%). There are more numbers with two or fewer components (69%) than numbers with three or more components (31%). For regular numbers and decimals, non-complex numbers (two or fewer components) are more frequent, while it is the opposite for dates, mostly because years often contain at least 3 components (e.g. two thousand seven). The distribution of numbers over speech types is very similar, except that impromptu speeches contain more regular numbers with three or more components (18% compared to 11% and 10%) and fewer dates with two or fewer components (3% compared to 9% and 10%).

Table 4 presents the overall distribution of renditions. For erroneous renditions (i.e. approximations, related and unrelated substitutions), the percentage of components that were incorrect is mentioned under the column ‘% of error’.

Table 4. Distribution of renditions

Category	#	Subcategory	#	% of error	#
Complete rendition	568				
		Interpreter’s correction	3		
Interpreter’s repetition	4				
Interpreter’s addition	7				
Approximation	9			25%	1
				33%	3
				50%	5
Related substitution	16	Phonological substitution	5	50%	1
				33%	2
				100%	2
		Syntactic substitution + Contextual substitution	1	25%	1
		Syntactic substitution + interpreter’s wrong correction	2	33%	1
				50%	1
		Syntactic substitution	7	25%	1
				50%	6
		Contextual substitution	1		
Unrelated substitution	28	Complete substitution	4	100%	5
		Complete substitution + interpreter’s wrong correction	1		
		Partial substitution	23	25%	1
				33%	4
				50%	14
				66%	2
				75%	1
				80%	1
Complete omission	97				

The categories ‘interpreter’s repetition’ and ‘interpreter’s addition’ cannot be considered errors as they do not automatically constitute mistakes and are therefore excluded from the analysis. The most frequent rendition type is ‘complete rendition’ (79%) followed by ‘complete omission’ (14%). It should be noted that the success rate in the corpus is substantially higher than in experimental studies and is close to the success rate of interpreters using visual support for numbers (Desmet et al. forthcoming). Besides complete renditions and omissions, 53 real errors are reported amounting to 7% of the cases. The most frequent type of mistake is unrelated substitution (28), and more specifically unrelated partial substitution with an error rate of 50% (14). Three categories of rendition are absent from the list: explicit approximation, partial and complete inversion. This is quite surprising as we would expect them to be fairly frequent in interpreting. The syntactic properties of numbers in one of the languages included in the corpus (Dutch) was indeed expected to trigger inversions. While additions are pretty rare (7), the mere fact that they exist is quite surprising. When having a closer look at these occurrences, it appears that they mostly concern additional information (e.g. a document’s publication date), stylistic expressions (‘This never happened in thousands of years’).

3.2 Sex differences in the rendition of numbers

The potential influence of the interpreter's sex on the rendition of numbers is the first part of this research project. Out of the 143 interpretations, 71 were performed by women and 72 by men. Male interpreters were exposed to 432 numbers, against 296 for female interpreters. Table 5 shows the distribution of categories of rendition between male and female interpreters.

Table 5. Categories of rendition for male and female interpreters

Category of rendition	Females		Males	
	#	%	#	%
Complete rendition total	**230**	**78%**	**338**	**79.7%**
Correction	*1*	*0.3%*	*2*	*0.5%*
Errors total	**21**	**7.1%**	**32**	**7.5%**
Approximation: total	**5**	**1.7%**	**3**	**0.7%**
25% of errors	*0*		*1*	*0.2%*
33% of errors	*2*	*0.7%*	*1*	*0.2%*
50% of errors	*3*	*1%*	*1*	*0.2%*
Related substitution: total	**9**	**3.1%**	**8**	**1.9%**
Phonological total	3	1%	3	0.7%
33% of errors	*1*	*0.3%*	*1*	*0.2%*
50% of errors	*1*	*0.3%*	*1*	*0.2%*
100% of errors	*1*	*0.3%*	*1*	*0.2%*
Syntactic total	6	2%	4	0.9%
25% of errors	*0*		*1*	*0.2%*
+ contextual	*1*	*0.3%*	*0*	
33% of errors + wrong correction	*0%*		*1*	*0.2%*
50% of errors	*5*	*1.7%*	*1*	*0.2%*
+ wrong correction	*0%*		*1*	*0.2%*
Contextual	0		1	0.2%
Unrelated substitution: total	**7**	**2.4%**	**21**	**5%**
Complete total	3	1%	2	0.5%
+ wrong correction	*0*		*1*	*0.2%*
Partial total	4	1.4%	19	4.5%
25% of errors	*0*		*1*	*0.2%*
33% of errors	*0*		*4*	*0.9%*
50% of errors	*2*	*0.7%*	*12*	*2.8%*
66% of errors	*1*	*0.3%*	*1*	*0.2%*
75% of errors	*1*	*0.3%*	*0*	

80% of errors	*0*		*1*	*0.2%*
Complete omission	**43**	**14.6%**	**54**	**12.7%**
Interpreter's addition	**2**		**5**	
Interpreter's repetition	**1**		**3**	

Given that males are exposed to more numbers than females (432 against 296), the type and complexity of numbers were compared between the two groups. The Chi-Square test shows that there is no statistically significant difference (X2=6.418, *p*=.268, effect size=0.09). The distribution of the type and complexity of numbers between male and female interpreters is presented in Table 6.

Table 6. Categories of number for male and female interpreters

Category	Females						Males					
	two or less		three or more		Subtotal		two or less		three or more		Subtotal	
	#	%	#	%	#	%	#	%	#	%	#	%
Numbers	176	59%	32	11%	208	70%	234	54%	49	11%	283	65%
Dates	25	8%	56	19%	81	27%	40	9%	84	20%	124	29%
Decimals	4	2%	3	1%	7	3%	20	5%	5	1%	25	6%
Total	205	69%	91	31%	296	100%	294	68%	138	32%	432	100%

In order to explore the first research question, a Chi-Square test was performed to assess the influence of sex on the rendition of numbers on five categories (complete omission, complete rendition, approximations, related and unrelated substitution). No significant difference was found (X^2=5.796, p=.215, r=0.09). Detailed results are found in Table 7.

Table 7: Sex differences in the rendition of numbers

	Females		Males		Chi-Square test		
	#	%	#	%	X^2	*p*	Effect size
Complete rendition	230	78%	338	79.7%	5.796	0.215	0.09
Approximation	5	1.7%	3	0.7%			
Related substitution	9	3.1%	8	1.9%			
Unrelated substitution	7	2.4%	21	5%			
Complete omission	43	14.6%	54	12.7%			

For the second research question, potential sex differences in the length of EVS for numbers were analysed with a Mann-Whitney U test at the interpretation level. The results show that there is no statistically significant sex differences (U=2538.500, p=.944, effect size=0.06). The results are presented in Table 8.

Table 8. Sex differences in number EVS

	N	Mean	M	SD	Mann-Whitney U test		
					U	p	Effect size
Number's EVS for females	71	2.96s	2.57s	1.34s	2538.500	.944	0.06
Number's EVS for males	72	2.85s	2.68s	1.14s			

3.3 Influence of predictors on the rendition of numbers

In the second part of this study, the influence of all identified predictors on the rendition of numbers is analysed: the interpreter's sex, the source and target languages, the source speech type (impromptu, read of mixed), the number type (dates, decimals and others) and complexity (two or fewer and three or more), the mean EVS, the number EVS, the source speaker's delivery rate, the source speaker average duration of silent pauses per minute and number of silent pauses per minute. The values for the continuous variables are presented in Table 9.

Table 9. Continuous predictors

Continuous variables	N	Mean	M	SD
Mean EVS	718	2.96s	2.82s	0.81s
Number's EVS	6213	2.74s	2.43s	1.50s
Source speaker's delivery rate	718	155w/m	154w/m	20.8w/m
Source speaker average duration of silent pauses per minute	718	10.4s	9.7s	3.1s
Source speaker number of silent pauses per minute	718	22.1	22.3	4.7

In order to assess the influence of predictors on the rendition of numbers, logistic mixed models are carried out on three categories of rendition: complete renditions, complete omissions and erroneous renditions (which include approximations, related substitutions and unrelated substitutions). The models were carried out on three categories only because sample sizes for each type of erroneous rendition are small. Details of the first model on the category "complete rendition" are found in Table 10 (N=568).

---

[3] The reduced sample size is due to the exclusion of the category 'complete omission'.

Table 10. Influence of predictors on complete renditions.

Predictors	β	Std. Error	z-value	Sig.
Significant predictors				
Interpreter’s language: French	-0.687	0.329	-2.091	.037
Interpreter’s language: English				
Number’s EVS	0.007	0.001	8.234	<.001
Mean EVS	-0.009	0.002	-5.044	<.001
Source speaker’s delivery rate	-0.022	0.007	-2.999	.003
Not significant predictors				
interpreter’s sex	0.095	0.233	0.408	.684
Interpreter’s language: Dutch	0.071	0.322	0.220	.826
Source speaker’s language: French	-0.412	0.352	-1.172	.241
Source speaker’s language: Dutch	-0.238	0.354	-0.672	.502
Source speech type: mixed	-0.295	0.442	-0.668	.504
Source speech type: read	-0.642	0.340	-1.890	.059
Number type	0.097	0.226	0.427	.670
Number complexity	-0.118	0.267	-0.442	.659
Source speaker average duration of silent pauses per minute	0.029	0.053	0.544	.587
Source speaker number of silent pauses per minute	0.002	0.029	0.070	.944

The marginal and conditional R^2 are both equal to 0.366. Four predictors have a significant p-value; the interpreter’s languages French and English (p=.037 and p<.001), the number’s EVS (p=<.001), the mean EVS (p=<.001) and the source speaker’s delivery rate (p=.003). French interpreters have fewer complete renditions compared to English interpreters. When the number’s EVS increases, the number of complete renditions increases. However, when the mean EVS and the source speaker’s delivery rate increase, the number of complete renditions decreases.

Table 11 presents the results of the same model for the category "erroneous renditions" (N=53).

Table 11. Influence of predictors on erroneous renditions.

Predictors	β	Std. Error	z-value	Sig.
Significant predictors				
Number's EVS	0.005	0.001	5.387	<.001
Source speech type: mixed	2.017	0.849	2.375	.018
Source speech type: read	2.215	0.798	2.775	.006
Source speech type: impromptu				
Not significant predictors				
interpreter's sex	0.093	0.343	0.272	.786
Interpreter's language: French	0.281	0.480	0.586	.558
Interpreter's language: Dutch	-0.208	0.470	-0.444	.657
Source speaker's language: French	0.453	0.549	0.824	.410
Source speaker's language: Dutch	0.044	0.520	0.085	.933
Number type	-0.041	0.310	-0.134	.894
Number complexity	0.203	0.357	0.569	.570
Source speaker average duration of silent pauses per minute	-0.043	0.077	-0.552	.581
Source speaker number of silent pauses per minute	0.079	0.048	1.625	.104
Source speaker's delivery rate	0.012	0.010	1.191	.234
Mean EVS	-0.002	0.003	-0.565	.572

The marginal and conditional R^2 are both equal to 0.313. Two predictors have a significant p-value: the number's EVS (p=<.001) and the source speech type (p=.018 and p=.006). When the number's EVS increases, the erroneous renditions increase. There are more erroneous renditions for read and mixed speeches compared to impromptu speeches.

Finally, the logistic mixed model was performed on complete omissions (N=97). Details are found in Table 12.

Table 12. Influence of predictors on complete omissions.

Predictors	β	Std. Error	z-value	Sig.
Significant predictors				
Mean EVS	0.006	0.002	3.360	.001
Source speaker's delivery rate	0.027	0.008	3.351	.001

Source speaker number of silent pauses per minute	-0.076	0.036	-2.131	.033
Not significant predictors				
interpreter’s sex	-0.071	0.282	-0.250	.803
Interpreter’s language: French	0.682	0.385	1.774	.076
Interpreter’s language: Dutch	0.014	0.397	0.036	.971
Source speaker’s language: French	-0.050	0.424	-0.118	.906
Source speaker’s language: Dutch	0.300	0.413	0.726	.468
Number type	0.036	0.255	0.141	.888
Number complexity	-0.383	0.308	-1.246	.213
Source speaker average duration of silent pauses per minute	-0.011	0.067	-0.157	.875
Source speech type: mixed	-0.972	0.500	-1.944	.052
Source speech type: read	-0.231	0.361	-0.638	.523

The marginal R2 is 0.191 and the conditional R2 is 0.237. The number’s EVS is not a predictor in this model given that there is no rendition in the interpretation and therefore no EVS. Three predictors have significant p-values: the mean EVS (p=.001), the source speaker’s delivery rate (p=.001) and the source speaker number of silent pauses per minute (p=.033). The number of complete omissions increases when the mean EVS and the source speaker’s delivery rate increases. When the source speaker’s number of silent pauses per minute increases, the number of complete omissions decreases.

The summary of the influence of predictors on each type of rendition is presented in Table 13.

Table 13. Summary of the influence of predictors

Predictors	Complete rendition	Erroneous rendition	Complete omission
Number's EVS	0.007	0.005	
Mean EVS	-0.009		0.006
Source speaker's delivery rate	-0.022		0.027
Source speaker number of silent pauses per minute			-0.076
Source speech type: mixed		2.017	
Source speech type: read		2.215	
Source speech type: impromptu	0		
Interpreter's language: French			
Interpreter's language: Dutch			
Source speaker's language: French	-0.687		
Source speaker's language: Dutch			
Source speaker's language: English	0		
Interpreter's sex			
Number type			
Number complexity			
Source speaker average duration of silent pauses per minute			

Table 13 shows that four predictors are not significant for all types of categories: interpreter's sex, source speaker average duration of silent pauses per minute and number type and complexity. The numbers' EVS has a positive influence on both predictors. The mean EVS and the source speaker's delivery rate have a negative influence on complete renditions and a positive influence on complete omissions. The source speaker number of silent pauses per minute has a negative influence on complete omissions only. Source speech type has an influence only erroneous renditions only and source speaker's language on complete renditions.

Given the significance of the number's EVS and its particular nature (no data for the category "complete omission"), its influence is analysed again on four categories: complete renditions, approximations, related and unrelated substitutions. A Kruskall-Wallis test is carried out to analyse the influence of the number EVS on the renditions and gave a significant p-value (p<.001), as well as a medium effect size (0.5). When looking at the actual values of EVS in each group, the EVS is longer for all types of erroneous renditions compared with complete rendition. Unrelated substitutions have the longest EVS (M=3.91s), followed by related substitution (M=3.23s) and approximations (M=2.69s). Detailed results are presented in Table 14.

Table 14. Influence of number's EVS on four categories of rendition

Number's EVS	N	Mean	M	SD	Kruskall-Wallis H test		
					X^2	*p*	Effect size
Complete rendition	568	2.62s	2.32s	1.41s	32.13	<.001	0.5
Approximation	9	3.59s	2.69s	1.97s			
Related substitution	16	3.61s	3.23s	1.87s			
Unrelated substitution	28	4.25s	3.91	1.77			

4. Discussion

This research project is one of the few corpus-based studies on the rendition of numbers in interpreting. Given that interpreters often receive help from their colleagues or have the written speech when dealing with numbers in real life, the frequency of errors and omissions in the data was much lower than in experimental designs (a total of 21% in our data compared to about 40% for experiments). The difference between corpus data and experiments could also be explained by the nature of the source speeches and the fact that interpreters at the European Parliament might be used to the type of numbers they generally deal with. In previous experiments, the most common type of rendition (besides complete rendition) is complete omission, followed by approximation (Desmet et al. forthcoming; Mazza 2001). Our data confirm that omission is the most frequent category after complete rendition, as it amounts to 14% of renditions. For the second most frequent category, our data show that unrelated partial substitutions are more frequent than approximations (with 3% of the renditions, compared to 1%). However, it has to be noted that our definition of approximations is more restrictive than in previous studies, therefore some studies might have included unrelated partial substitutions under the category 'approximation'.

When it comes to the influence of sex on the rendition of numbers, the regression and the Kruskall-Wallis H test did not yield significant results. Moreover, no sex differences were found in the length of the Ear-Voice span for numbers. The analyses of the influence of other predictors on the rendition of numbers seem to indicate that several predictors have significant p-values. Timarová et al. (2014) suggest that professional interpreters perform better than students because professionals keep a shorter EVS. Our results for the mean Ear-Voice Span (measured at the world level for a whole speech) partly confirm this hypothesis as they show that the EVS is indeed shorter for complete renditions and longer for complete omissions, but is not significant for erroneous renditions. However, the Ear-Voice span for numbers (measured for each number but absent for complete omissions) is longer for erroneous renditions as well as for complete renditions. Setton (1999) observes that errors typically occur when the EVS is more than 3–4 seconds. This is confirmed by the Mann-Whitney U test, which shows that the EVS is significantly longer in case of erroneous renditions compared to complete renditions ($p<.001$ and effect size=0.5), and most specifically longest for unrelated substitutions (mean=4.25s), followed by related substitutions (mean=3.61s) and approximations (mean=3.59s). This suggests that the more accurate the target number is, the shorter the EVS. It was suggested that the source speech type might have an impact on the rendition of numbers, as a read speech would potentially mean that interpreters have the speech in front of them and can therefore read the numbers as they are being said, or received the help of their colleagues. In our results, this variable is only significant for erroneous renditions and shows that the number of erroneous renditions is higher for mixed and read speeches, compared to impromptu speeches. The previous hypothesis is therefore not confirmed by these data. It is, however, important to note that the information about source speech type is sometimes incomplete and to the fact that some speeches were partly read and partly impromptu (labelled as 'mixed'). The source speaker's delivery rate was also mentioned as increasing the error rates (Korpal 2016). This is partly confirmed by our data which show that, as the source speech delivery rate increases, the number of complete omissions increases, but the number of complete renditions decreases. However, this predictor has no influence on the category "erroneous renditions". When it comes to the source speaker's silent pauses, our data show that the number of silent pauses per minute is only significant for complete omissions. As the number of silent pauses increases, the number of omission decreases, possibly meaning that the input is less dense and the interpreter has more resources to dedicate to numbers.

Pinochi (2009) suggests that differences in number syntax between source and target language can increase the error rate. Despite the inclusion of languages with different syntax (Dutch versus French and English), this hypothesis was not confirmed as target languages did not yield significant p-values, as well as Dutch as a source language. French and English as source languages are only significant for complete renditions, where interpreters from French produce fewer complete renditions than interpreters from English. This might due to the somewhat complex nature of numbers in French compared to English (e.g. "ninety" literally translates into "eighty-twenty-ten" in French). Previous experiments also show that most errors seem to happen with large numbers and decimals (Desmet et al. forthcoming). Our data did not confirm this as no significant p-values were found for the influence of the type nor the complexity of numbers.

5. Conclusion

Given the female advantage in several processes involved in simultaneous interpreting, more specifically short-memory tasks, this research project aimed at determining whether sex differences can be found in the rendition of numbers by interpreters at the European Parliament. The first hypothesis stipulates that women have more cognitive resources available when dealing with numbers and therefore render more numbers than men. Indeed, omissions and errors are believed to be caused by a lack of attentional resources required to perform the task. Our data refute this hypothesis as no sex differences were found in the rendition of numbers, nor in the type of erroneous renditions. The second hypothesis concerns potential sex differences in the EVS associated with numbers. It was assumed that men have fewer cognitive resources and memory capacity available when dealing with numbers, and therefore resort to one of the strategies applied by professionals, i.e. shortening the EVS, more often than women. This hypothesis was also refuted by our data as no sex differences were found in EVS lengths. To conclude, our data clearly show that sex plays no significant role in the rendition of numbers.

Secondly, this research project also aimed at identifying relevant predictors of the rendition of numbers. Several predictors seem to be significant for several types of renditions: the Ear-Voice Span (measured at world level and for each number), the source speech type (read, mixed and impromptu speeches), the source speech delivery rate and number of silent pauses per minute as well as the source languages French and English. This study tends to show that the shorter the EVS, the more accurate the rendition of numbers. These results confirm the practices already used by most professionals: shortening the EVS is a useful strategy when dealing with numbers. Trainers would therefore be well advised to include this strategy in their training programme. The results also show that complete omissions are more numerous when the input is faster and the number of silent pauses lower. The source speaker's delivery rate also reduces the number of complete renditions, but had no influence on erroneous renditions. This might indicate that, faced with a fast or dense speech, interpreters prefer to omit than to make a mistake. Our data suggest that erroneous renditions are more numerous with mixed and read speeches, showing that either interpreters do not have the speech in advance, or if they do, they still make more mistakes than for impromptu speeches. To conclude, it seems that two types of predictors are particularly significant for numbers: the length of EVS and predictors linked to the source speech.

An additional objective of this research project is to observe the rendition of numbers in a real environment. As expected, the occurrence of numbers, and consequently of errors, was lower for corpus data compared with experimental data but some interesting findings were reached. Some findings from experiments were confirmed: the influence of a shorter EVS on the rendition of numbers, as well as the fact that omissions are more frequent than erroneous renditions. However, the influence of some previously identified predictors was refuted: the type and complexity of the number and languages with different word orders. These discrepancies might also be due to the small sample sizes in the data. Therefore, more corpus data are needed in order to analyse real-life errors and their predictors in more details.

Chapter 3: Disfluencies in simultaneous interpreting, a corpus-based study with special reference to sex

Abstract

This paper reports on a study on disfluencies (filled and silent pauses, false starts and lengthenings) in simultaneous interpreting carried out on corpus data drawn from the European Parliament Interpreting Corpus Ghent (EPICG). The aim of this research project is to analyse the influence of the interpreter's sex alongside sixteen predictors in the production of disfluencies. Simultaneous interpreting is a cognitively demanding task for which interpreters often experience high cognitive load. Studies show that females seem to outperform males on verbal tasks (Herlitz et al. 1997; Loonstra et al. 2001; Maitland et al. 2004; Hirnstein et al. 2014 inter alia) and on memory tasks (Trahan & Quintana 1990; Kimura & Seal 2003 inter alia). Despite these differences and the potential impact the existence or absence of sex differences could have on research and training, few interpreting studies have sex as a control variable (Cecot 2001; Mason 2008; Baes 2012; Defrancq 2013; Magnifico & Defrancq 2016; 2017). This research project therefore attempts to determine whether sex differences have an influence on disfluencies. If women perform better at verbal tasks and need to dedicate less energy to the concurrent cognitive processes involved in simultaneous interpreting, they are less likely to be affected by cognitive load. Since disfluencies are believed to be the consequence of cognitive load, we can assume that female interpreters produce fewer disfluencies than male interpreters. The data sample consists of 180 source texts and 180 interpretations in six language pairs (from and into English, French and Dutch) and is based on a parallel acoustic aligned and time-tagged corpus.

Results partly confirm the hypothesis as they show that male interpreters produce more lengthenings and false starts and have longer duration of silent pauses than female interpreters. The analyses also reveals several significant predictors of disfluencies, such as the Ear-Voice Span and the input delivery rate.

Keywords: simultaneous interpreting, disfluencies, corpus, sex differences.

1. Introduction

The aim of this paper is to analyse the influence of the interpreter's sex1, as well as other predictors, in the production of disfluencies (filled and silent pauses, restarted sentences, words with elongated pronunciations, etc.) in simultaneous interpreting (SI). Few researchers have analysed the influence of sex in interpreting studies. Yet, there are reasons to believe that sex is a significant variable for interpreting. While the existence of sex differences in cognitive skills in general is uncertain, sex differences have been found for specific tasks related to the production, memory and analysis efforts in interpreting (Hyde & Linn 1988; Maitland et al. 2004; Aerts 2013; Hirnstein et al. 2014 inter alia). The few studies focusing on sex differences in interpreting have concluded that differences exist (Cecot 2001; Mason 2008; Baes 2012; Magnifico & Defrancq 2016 inter alia). Studies on cognitive sex differences are useful for several reasons. First, they can supplement and nuance studies on gender differences either by giving some explanations for the differences found or by suggesting that gender differences are societal, cultural or educational. Second, they can give researchers and trainers a new perspective on their work by encouraging them to consider sex as a significant variable. Halpern et al. (2007) suggest that we can use the knowledge acquired through studies on sex differences to teach female and male students ways to solve problems that correspond to their most efficient cognitive process, which has a positive impact on their performance. Third, it is a well-known fact that the interpreting profession is dominated by women (Lim 2005; Ryan 2015; Hickey 2018). The reasons behind this phenomenon are still relatively unclear but Miller & Halpern (2014) suggest that individuals' relative cognitive strengths are important to career and educational choices. Studies on cognitive sex differences could therefore help explain this phenomenon, while potentially eliminating stereotypes. Studies on sex differences in verbal tasks showing a female advantage might have an impact on men's decisions, as men might not choose interpreting based on their impression of a female superiority in interpreting. This perception could also have an impact on their performance, given that people's performances appear weaker when they are told that the other sex performs better at the task (Spencer et al. 1999).

Disfluencies are believed to be triggered by cognitive load, which, in turn is believed to be determined both by the cognitive demands imposed by "the individual concurrent tasks" (Seeber 2011: 187) involved in interpreting and by the available

---

[1] Throughout this paper, the authors refer to the interpreter's sex and not the interpreter's gender. While the expression 'gender' is commonly used, notably in reference to gender differences, it generally refers to an individual's self-conception and role within society. In fact gender differences studies tend to focus on communicative and linguistic differences (Coates 1993; Chambers & Trudgill 1998) while the present study analyzes the cognitive aspects and therefore only takes the subjects' biological sex into account and makes no assumptions on their gender.

resources to carry out those tasks. Whereas most of the literature focuses on the tasks and their effect on cognitive load and properties of interpreting (Gerver 1969; Barik 1975; Gile 1995; Seeber 2011; 2013 inter alia), this study will also focus on the resources. Females are generally, but not systematically, found to perform better than males in carrying out tasks related to SI (see Section 1.3), which begs the question whether female interpreters are also more resourceful than male interpreters to cope with cognitive load. A quantitative study on disfluencies in female and male interpreting could shed some light on the matter, as disfluencies are believed to be the result of cognitive overload, i.e. a situation in which the available resources cannot meet the demands. The main hypothesis of this study is therefore that female interpreters produce fewer disfluencies than male interpreters. However, female and male perform in particular circumstances that inevitably also have a bearing on their performance. We will therefore also include a series of background variables related to the circumstances in our analysis.

This research project is based on a parallel acoustic aligned and time-tagged sub-corpus of the European Parliament Interpreting Corpus Ghent (EPICG). The data used for this study comprise 30 source speeches and their interpretations for six language pairs (English-French, French-English, English-Dutch, Dutch-English, Dutch-French and French-Dutch) for a total of 14 hours of interpreted speech and 108245 interpreted words. The following types of disfluencies were selected: filled and silent pauses, false starts and lengthenings. The corpus offers a wide variety of speakers, speeches and languages. It was therefore decided to also look at the variables that potentially influence the production of disfluencies. The following variables are included in the data: interpreter's sex, source and target languages, Ear-Voice Span, source speaker/interpreter speaking time ratio, source speaker's and interpreter's delivery rate, number of filled pauses, lengthenings, false starts, silent pauses and the total duration of silent pause.

In the first section, this article explains the concepts of cognitive load associated with disfluencies and with simultaneous interpreting. The literature on sex differences for several cognitive skills relevant to simultaneous interpreting is also covered. In the second section, the corpus is presented, as well as the methodology used to identify and measure the disfluencies and predictors. Descriptive statistics and the results of the statistical analyses for sex differences and of the multiple regressions are presented in the third section. Finally, the results are discussed in the last section.

1.1 Cognitive load and disfluencies

Disfluencies are defined as "phenomena that interrupt the flow of speech and do not add propositional content to an utterance" (Fox Tree 1995: 709). There is considerable variety in the literature on which items are to be considered disfluencies as well as on their classification (Shriberg 1994). In most studies, a distinction is made between repair and non-repair categories. A repair is an occurrence of interrupted speech combined with an attempt at producing an alternative for a previously articulated segment (Levelt 1983). Shriberg's (1994) own classification in single-token disfluencies and structured disfluencies, for instance, partly rests on that distinction. Except for the lexical false starts, single-token disfluencies do not involve a repair: filled pauses (uh, uhm), intra-word pauses and hesitation-related lengthening of phones. On the other hand, structured disfluencies include typical cases of repair: repetitions (of one or more words), deletions (cases where a segment is discarded and the structure is started anew), substitutions (cases where segments are replaced with others in the same structure); insertions (when a segment is repeated with an additional item). As explained in the methodology, we will select a subset of the most easily identifiable disfluencies for this study, i.e. false starts, filled pauses, silent pauses (intra-word but also between words) and lengthenings. The latter three pertain to the non-repair type, while the first is a repair. Silent pauses are a complex type of disfluencies as they are not always a disfluency, they can also be strategically used as a rhetorical device, or to maintain the prosodic structure of an utterance (MacGregor et al. 2010).

An alternative classification based on the causes of disfluencies is put forward by Gósy (2007), who differentiates between two major groups of speech disfluencies: (1) disfluencies rooted in uncertainty (UDs) such as hesitations, fillers, repetition, restarts, lengthening and pauses within the word and (2) error-type disfluencies (ETDs) such as Freudian slips, grammatical errors, contamination, false word activation, tip of the tongue, ordering problems and slips2.

Disfluencies are frequent in speech: averaging across a number of studies, and excluding silent hesitations, it has been estimated that disfluency in spontaneous speech affects about 6 per 100 words (Bortfeld et al. 2001). Demographic variables seem to influence the frequency of disfluencies: Bortfeld et al. (2001) and Shriberg (1996) both report that men produce more disfluencies than women. Engelhart et al. (2010), however, do not find significant differences between the sexes. The frequency of disfluencies also

---

2 A lengthening is defined as a marked prolongation of one or more phones (often limited to one syllable), resulting in above-average syllable and word duration (Betz et al. 2015; Betz & Wagner 2016). A false start is an utterance that is aborted and then restarted. Lexical identical repetitions happen when the speaker produces the same lexical form multiple times in a row.

appears to increase with age (Shewan & Henderson 1988). Demographic variables are interesting because they consider the capacity dimension of cognitive load. Cognitive load rests on the idea that the human working memory has only limited capacity, which prevents it from performing several tasks simultaneously at the same speed and the same level of efficiency as when the tasks are performed separately (Welford 1952; Broadbent 1958). Working memory also has limited capacity for storing the information that is necessary to perform the tasks. If capacity does not meet demands, disfluencies tend to appear or increase. Differences on the sex or age dimension could be due to capacity differentials. This study will explore the former dimension.

Disfluencies can be triggered by a variety of situations that impose increased demands: new information (Clark & Fox Tree 2002; Arnold et al. 2003), heavy constituents (Swerts 1998; Arnold et al. 2000; Watanabe et al. 2008) or long sentences (Oviatt 1995; Shriberg 1996). Boomer (1965) and Shriberg (1996) found more disfluencies near the beginnings of turns or sentences, where planning effort is presumably higher. The topic or domain of a conversation is another characteristic that may cause the planning load of utterances to vary. In one study by Schachter et al. (1991), it was found that social science lectures contained more fillers than hard science lectures, while humanities lectures contained most.

While disfluencies are seen as the consequence of cognitive load, they can also have a positive influence on the listener. The small number of studies that have investigated the effect of disfluencies on comprehension show that under specific circumstances, disfluencies can help the listener by helping the identification of upcoming words (Howell & Young 1991; Fox Tree & Schrock 1999; Brennan et al. 2001; Fox Tree 2001; Arnold et al. 2003, Arnold et al. 2004; Ferreira et al. 2004). In specific contexts, disfluencies are used as communicative cues as they signal troubles in delivery to the listener and aid in comprehending the intended grammatical and semantic structure (Betz et al. 2015). If a speaker hesitates, the listener expects the speaker to say a difficult word, a non-predictable word because the listener assumes a higher cognitive effort and will therefore focus more (Kutas & Hillyard 1984; Corley et al. 2007; Collard et al. 2008, Macgregor et al. 2009). Corley & Hartsuiker (2011) find that filled pauses (uh, um, and the like) serve to signal upcoming delays in a way which informs listeners' reactions and facilitate word recognition. Fox Tree (2001) found that both English and Dutch listeners are faster to identify a target word in a carrier sentence when it follows an uh in comparison to a control condition without the uh. It is also possible that the benefits for perception emerge from the fact that disfluencies like uh and um, and any silent pauses preceding or following the filler, considerably delay target word onset themselves. Anecdotal evidence suggests that speakers who are difficult to follow will be more easily understood when they speak more slowly (and therefore pause more).

1.2 Cognitive load in simultaneous interpreting

Simultaneous interpreting is considered a cognitively demanding task (Gerver 1969; Barik 1975; Gile 1995; Seeber 2011; 2013 inter alia). It includes following cognitive processes, such as speech comprehension and production, memory, attention/resource allocation and the monitoring of simultaneous operations (Klaudy 2004). Interpreters are required to deliver their speech in a fluent and efficient way, making speech production and fluency key skills for simultaneous interpreters. The speech production system mobilizes the mental lexicon, knowledge of the outside world, and a syllabary (Levelt 1999), while being kept in check by a self-monitoring mechanism (Postma 2000) that inspects one's own speech and takes appropriate action when errors are made (Hartsuiker & Kolk 2001). Disfluencies can also be seen as a window into the cognitive processes of speech planning (Goldman-Eisler 1958; Nooteboom 1969; MacKay 1973; Garrett 1975; Shattuck-Hufnagel 1982; Levelt 1989; Bock & Levelt 1994). Disfluencies are connected to the problem of lexical access in the form of false word activations, and the disharmony between lexical access and articulatory planning in the form of prolongations, and restarts (Tóth 2011).

SI can also be seen as a special case of speech production in noise (Tóth 2011). Indeed the condition of a task performed in noisy environments and interpreting is similar: the need to divide attention. Interpreters must divide their attention between listening to the input utterance and rendering the translation and therefore work under high cognitive pressure. Interpreting is a good example of a process where demands compete for cognitive capacities and is believed to be linked to the ability to manage competing demands on limited cognitive resources (Liu et al. 2004). Gile (1997) and Seeber (2011) have designed capacity-demand models that represent the interpreting process as a 'cognitive management problem' where interpreters need to find a balance between available resources and demands imposed by the different sub-tasks of interpreting. When the processing load exceeds the processing capacity, the interpreter will most likely produce errors and omissions, as well as disfluencies. Gile (1995) confirms that restarts and lengthenings, two types of disfluencies occurring in the output of simultaneous interpreters, might be partly explained by the mental energy required for the task.

Disfluencies are therefore likely to be particularly frequent in interpreting. Several studies show that speech produced under high cognitive load typically presents more disfluencies (Jameson et al. 2009; Yap 2012; Schuller & Batliner 2013). Christodoulides & Lenglet (2014) found more filled pauses, false starts, repetitions and deletions as well as longer pauses for interpreting compared to reading (in total, 9.8% of the tokens were disfluent in SI, compared to 0.4% in Reading). However, their study is based on only two experimental subjects. Other studies, carried out on larger populations show that interpreters produce longer silent pauses compared to the source speech (Tissi 2000; Cecot 2001; Ahrens 2005; Christodoulides 2013), more false starts (Pöchhacker 1995; Tissi

2000), more numerous vowel and consonant lengthening (Tissi 2000) and more filled pauses (Plevoets & Defrancq 2016; 2018). Mead (2000) has also shown that students produced more filled pauses when interpreting into their B language than into their A language, because the former is more cognitively demanding than the latter. Bakti (2009) found that restarts and grammatical errors are the most frequently occurring disfluency in a corpus of simultaneous interpreters working from English into Hungarian.

Moreover, several factors are known to increase the cognitive load and are therefore likely to have an influence on the frequency of disfluencies. Seeber (2011) lists two input features: delivery rate and language pair. While delivery rate is reported not to increase cognitive load significantly in the comprehension of spontaneous speech (Voor & Miller 1965), it has been proven to influence interpreters' performances considerably (Gerver 1969; Pio 2003). Seeber (2011) suggests that simultaneous interpreting of Subject-Object-Verb (e.g. German and Dutch) into Subject –Verb-Object (e.g. French and English) structures generates more cognitive load than interpreting SVO into SVO structures. This is mostly due to the additional cognitive processing triggered by the strategies (i.e. waiting, stalling and chunking) applied by interpreters when dealing with syntactic asymmetry between source and target languages. Gile (2008) highlights source text sentence length as one of the factors increasing cognitive load in interpreters, adding that length as such is probably not a factor, but the syntactic complexity that comes with it in most cases. He also points to the lexical density as one of the prime determinants in cognitive load in interpreting. Chmiel & Mazur (2013) report that in an experiment on sight translation performed by trainee interpreters long sentences receive longer fixation times, indicating an increased cognitive load in the interpreter.

Research into the factors of cognitive load in interpreting has been predominantly source-oriented but it is reasonable to assume that the cognitive load of an interpreter is also influenced by his or her attempts to produce a target text under high cognitive load. In two studies on cognitive load and filled pauses, Plevoets and Defrancq (2016; 2018) found that both high source text delivery rates and high source and target text lexical density triggered more filled pauses in interpreters. High formulaicity has a significant negative effect on the occurrence of filled pauses. This suggests that the triggers of disfluencies are chiefly lexical. Since disfluencies are assumed to be influenced by the way interpreters divide their attention between different efforts (Gile 1995), we can assume that the length of the Ear-Voice Span (i.e. the lag time between the source speaker and the interpreter) will also have an impact on disfluencies (Tóth 2011). Gile (1995) considers Ear-Voice Span (EVS) a possible indicator of how the various efforts relate to each other. Indeed, a long EVS might mean that the interpreter prioritizes the listening effort over the production effort, while a short EVS potentially means that the interpreter is saving memory capacity. The type of disfluency can also be influenced by the type of effort as Setton (1999) suggests that long silent pauses mean that the attention is dedicated to the listening task while long filled pauses mean that the attention is almost entirely dedicated to the production task.

It is vital, in a study exploring sex as a predictor for disfluencies in simultaneous interpreting to consider a wide variety of contextual variables, such as the ones discussed in this section. As repeatedly stated in the literature (Diriker 2004; Duflou 2016), simultaneous interpreting is a situated linguistic performance that cannot be dissociated from the context in which it takes place. A study on the effect of sex could very well yield completely misleading results if contextual variables were not taken into account.

1.3 Sex differences in cognitive skills

Sex differences in cognitive skills are a sensitive and controversial topic that needs to be handled with care. Meta-analyses, which aim at determining whether sex differences exist as a whole or in global categories (mathematical, verbal and spatial among others), tend to indicate that males and females are much more similar than they are different and that sex differences are often exaggerated (Hyde & Linn 1988; Hyde 2005; Miller & Halpern 2014). For example, Hyde & Linn's (1988) meta-analysis on verbal abilities includes various tasks (spelling, reading, writing and vocabulary) and concludes that there is no scientific proof that allows claiming that women and men have different verbal abilities. This being said, several individual studies found sex differences for specific tasks and Hyde & Linn (1988) recognize that females do score significantly higher at one particular task: speech production. Accordingly, the aim of this research project is neither to prove that cognitive sex differences exist in general and outside of interpreting nor to generalize the findings to the rest of the population. This research project aims at exploring the influence of a rarely analyzed predictor (sex), alongside other know predictors, on disfluencies during one specific task (simultaneous interpreting).

As mentioned before, studies found sex differences in specific tasks which appear to be relevant for simultaneous interpreting. Besides having better speech production abilities (Hyde & Linn 1988), women have been found to have greater verbal fluency, i.e. the ability to retrieve specific information within restricted search and time parameters, for example the ability to generate words beginning with a single letter in one minute (Herlitz et al. 1997; Loonstra et al. 2001; Maitland et al. 2004; Hirnstein et al. 2014). A female advantage in generating synonyms has also been found (Hines 1990). Aerts (2013) found that women display a larger sensitivity to the phonemic contrasts during auditory phoneme discrimination and showed more differentiation in real word-pseudoword dissociation. Studies have also found faster processing speed (the speed at which a person can understand and react to the information they receive) in women (Keith & Reynolds 2008) and a female advantage in both pre-lexical and lexical processing was found (Majeres 1999), as well as for perceptual speed (the ability to compare or recognize items) (Born et al. 1987; Hedges & Nowell 1995). In other words, women could have an advantage for the listening and analysis effort in SI. For the memory effort, evidence for a female advantage in episodic and some aspects of semantic memory has been found (Herlitz et al. 1999; Kramer 1997; Maitland et al. 2004). Women also tend to perform better than men on immediate and delayed free recall and on recognition tasks with verbal and visual components (Trahan & Quintana 1990; Kimura & Seal 2003). However, Harness et al. (2008) report higher scores for males in a study on recall combined with a distraction task carried out on students, which could be relevant to our purposes if simultaneous interpreting is considered a language production task with the incoming speech as a distractor. Studies report that females outperform males in the Rey Auditory Verbal

Learning Test (free recall of two lists of nouns, which aims at evaluating short-term auditory-verbal memory, retroactive, and proactive interference, retention of information among others) and the Verbal Paired Associates test (immediate and delayed recall of word pairs, aimed at evaluating explicit episodic memory performance) (Bolla-Wilson & Bleecker 1986; Gale et al. 2007).

When it comes to the production of disfluencies, females have been found to suffer less frequently from clinical disfluency (or stuttering) (Guyette & Baumgartner 1988; Yairi & Ambrose 1992). Shriberg (1996) finds that men produced more fillers than women. Similarly, Bortfeld et al. (2001) show that men present higher rates of disfluencies overall (6.80 to 5.12 per 100 words), which is mainly due to higher rates of fillers and repeats. Men produced slightly but not significantly higher rates of restarts than women. Engelhart et al. (2010), however, do not find significant differences between the sexes.

As far as sex effects on simultaneous interpreting are concerned, only Cecot's (2001) experimental study seems to have taken them into account. She finds that females use more filled pauses, whereas males use more unfilled pauses, and that men's unfilled pauses last longer than women's. There is some evidence that female and male interpreters cope differently with pragmatic challenges in interpreting: Mason (2008) finds more omissions for males in consecutive interpreting in the courtroom and suggests that omission patterns are determined by gendered social behaviour. Magnifico & Defrancq (2016; 2017) show that female interpreters use more hedges and downtone fewer unmitigated face-threatening acts than male interpreters at the European Parliament and, interestingly, that female interpreter self-repair more often than male interpreters (Magnifico & Defrancq submitted).

1.4 Research question and hypotheses

The main research question of this study is what determines differences in the production of disfluencies between individual interpretations. Overall, the cognitively challenging nature of interpreting is expected to lead to more disfluencies in interpreting than in normal speech. Sex as a predictor for disfluencies will be explored, although the meta-literature on differences in linguistic performance and fluency between the sexes is generally inconclusive. Accounts of a female advantage in some linguistic tasks, production tasks in particular, do, however, exist and outnumber accounts of a male advantage. Therefore, our main hypothesis will be that female interpreting will present a lower frequency of disfluencies than male interpreting. To do justice to the complex nature of the interpreting task and to the situated nature of the interpreting activity, a list of potential contextual predictors was also included in the study. Some of them have been identified in previous research as increasing cognitive load in interpreters and are therefore expected to have an effect on the frequency of disfluencies in interpreting: speaker's delivery rate, EVS, language pair. The mutual, potentially reinforcing effect of disfluencies on other disfluencies will also be studied.

2. Methodology

2.1 A corpus-based approach

The present study is corpus-based, i.e. based on naturalistic data produced in a real-life environment by professionals and have therefore the potential to reinforce the empirical foundations of interpreting research (Shlesinger 1998). Thanks to new technologies that simplify the compilation, transcription and analysis of data, as well as to the availability of interpreting data online, corpus-based interpreting studies (CIS) are gaining in popularity. Back in the early 2000s, the University of Bologna compiled the first publicly accessible simultaneous interpreting corpus from data from the European Parliament, the Parliament Interpreting Corpus (EPIC) (Bendazzoli & Sandrelli 2005). Several universities have followed suit building their own interpreting corpora : CoSi (consecutive and simultaneous interpreting) and DiK (dialog interpreting in public service settings) at the University of Hamburg, EPICG (European Parliament Interpreting Corpus Ghent) at Ghent University, and others at the universities of Rome, Trieste, Posnan, Louvain-la-Neuve and Saarbrücken.

The data used for this study are drawn from EPICG, one of the corpora based on audiovisual recordings of speeches and interpretations at plenary sessions of the European Parliament. Speeches at the European Parliament are generally very short (1 to 6 minutes) and the working conditions only reflect the institutionalized context, which makes the corpus admittedly less representative for the interpreting activity at large. Therefore, the conclusions drawn by data coming from this context do not necessarily apply to all types of interpreting. On the other hand, interpreters working for the plenary sessions of the European Parliament are generally experienced and have undergone an accreditation test, which ensures a baseline interpreting quality in the data. Corpus-based studies are sometimes criticized because they consist of samples taken out of the context in which they occurred (Diriker 2004; Duflou 2016). It is therefore essential to provide metadata to give more information about the context (Burnard 2002). Metadata on the speaker (political group and function, age, sex) and the speech (topic, time of the day, delivery type) were added to the corpus based on information provided through the European Parliament's website.

The data used for this study are collected from a sub-corpus of EPIC Ghent which comprises 30 source speeches and their interpretations for six language pairs (English-French, French-English, English-Dutch, Dutch-English, Dutch-French and French-Dutch). These data offer a wide diversity of topics, speakers and interpreters. Source and target texts were transcribed according to the Valibel instructions (Bachy et al. 2007) and acoustically aligned on the basis of pauses with the transcription tool EXMARaLDA Partitur-Editor. More information on the compilation process, including transcription conventions and annotations can be found in Bernardini et al. (2018). One important aspect of EPICG is the availability of oscillograms for source speech and interpretation. These obviously facilitate the identification of disfluencies such as pauses. In total, the corpus comprises more than 14 hours of interpreted speech and a total word count of 108 245 interpreted words.

For each combination of source and target language, a balanced set of 15 male and 15 female interpretations was aimed at[3]. Given that the interpreters in the corpus are anonymous, the authors were faced with two methodological challenges: 1) sex had to be determined on the basis of the recorded voices only; 2) there is a possibility that some interpreters occur more than once in the database. In order to solve the first challenge, the sex of each interpreter was determined separately by both authors and an independent reviewer, which is a reliable method according to Lass & Puffenberg (1971). Indeed human listeners are able to identify speaker sex with an accuracy of over 95%. The process for our corpus yielded an inter-rater agreement of 99.4%, with the three assessors diverging on only one interpretation. It was concluded that the disagreement came from a human encoding mistake and the three assessors finally agreed on all interpretations. In order to complement the human identification process, a speaker diarization software LIUM_SpkDiarization (Rouvier et al. 2013) was also used. After several necessary adaptations to the software and the corpus's audio data, the human-machine agreement reached 95.6%. However two human assessors disagreed with the sex identified differently by the software and agreed with the human identification. The latter was therefore chosen as the reference. The software's lack of accuracy can be explained by the fact that it is optimized for radio and TV shows and it cannot guarantee the same level of performance for other types of recordings. Moreover, the diversity of languages and speakers and the complexity of the audio data (several speakers take the floor simultaneously and the quality is not always optimal) make the task more challenging for the software.

Additional steps were undertaken in order to tackle the second challenge and ensure a representative and diverse set of interpreters. First, the study analyses languages that are sufficiently common to guarantee that they are covered by a large number of interpreters (English, French and Dutch). Second, speeches were randomly sampled from the European Parliament's website over a 6-year period in order to reduce the risk of having the same teams in the interpreting booths. In the final dataset, 93 interpretations are drawn from 21 different dates in 2008, 47 interpretations from 17 different dates in 2009, 14 from 11 different dates in 2010, 9 from 9 different dates in 2011 and 16 from 4 different dates in 2013 and 1 in 2014. As a rule of thumb, a maximum of 3

---

[3] According to the estimations of the Organisation Intersex International Europe, there is a probability that one percent of the interpreters in the corpus are intersex. Unfortunately there is no possibility of knowing whether interpreters in the corpus are intersex. While the authors are conscious of this possibility, they consider that the size of the corpus means that this low probability does not obviate much of the discussion.

interpretations from the same interpreter for each target language and sex is considered as acceptable by the authors.

While human listeners seem to be more successful than the software at identifying sex, the identification of speakers for such a diversified corpus appears to be quite unreliable when performed by human listeners. Therefore, LIUM_SpkDiarization was used for the identification of identical interpreters and the results were rated as reliable by two human assessors. However, given the difficulty for these raters to identify the interpreters, no precise human-machine agreement could be determined. Table 1 therefore the results of the LIUM_SpkDiarization's analysis only.

Table 1. Identification of identical interpreters by LIUM_SpkDiarization

Sex	Language	Number of interpreters identified		Total number of unique interpreters
		Twice	Three times	
Females	French	3	0	27
	English	3	1	25
	Dutch	4	0	26
Males	French	3	0	27
	English	2	2	24
	Dutch	2	0	28

While in theory the presence of identical interpreters in the data set is a violation of the assumption of independence of observations within each group (males and females), the effects are likely to be very limited given that for each sub-corpus the ratio interpreters/interpretations is at least 0.80 and no interpreter occurs more than three times in the database.

2.2 Identification and measurement of disfluencies

For our research, the following types of disfluencies were selected: lengthenings, filled pauses, false starts and silent pauses. These disfluencies are relatively easy to identify and were frequently observed in the corpus. The task of identifying silent pauses and lengthenings was simplified by the presence of an oscillogram where these two types of disfluencies produce perceptible patterns. Where one disfluency occurred right after another or where there were several disfluencies in a row, these were coded as separate occurrences. No distinction was made between the different types of filled pauses (uhm, uh, hum and euh). False starts include truncated words ("the pre/ president") no matter whether the same word is repeated or if another word is uttered ("the pre/ chairman"). Silent pauses were included if they lasted at least 0.2 seconds and two types of measurements were taken into account: their frequency (i.e. the total number of silent pauses in one speech) and their total duration (i.e. the total duration silent pauses account for in a whole speech, contrary to the duration of one single silent pause). It is important to mention that all pauses were taken into account, not only intra-word pauses. Therefore, when analysing the results, we will make a distinction between the articulated disfluencies (i.e. lengthenings, filled pauses and false starts) and silent pauses as the latter do not necessarily constitute a disfluency.

While disfluencies were first manually annotated in the transcript, they were counted by a tailor-made script. Their frequency was determined in two ways: on the one hand, we performed a normalization per minute on the text level: the number of occurrences of a particular disfluency is divided by the total duration of the interpretation they occur in, and is then multiplied by 60 in order to obtain a number of disfluencies per minute. This measurement provides an average frequency of disfluencies in a whole speech. Secondly, since disfluencies (and other predictors) can vary highly throughout a speech, source speeches and interpretations were also divided into segments. As a rule of thumb, we determined segments of ca. 10 seconds, but as we wanted to avoid segment boundaries splitting up articulated portions of the acoustic signal, segment length varies. To avoid distortions, only segments of 8 to 12 seconds were included in the study (97.9% of the total number of segments). For each segment, the frequency of disfluencies is determined manually. For a clearer representation, these data were also normalized (each data was divided by the actual segment duration and multiplied by 10 seconds).

The distributions of all disfluencies being rightly skewed, the one-tailed non-parametric Mann-Whitney U-test was chosen to check for potential sex differences in each type of disfluency per interpretation. As a reminder, if the p-value associated with the Mann-Whitney U test is below the significance level (0.05), the difference between men and women is considered as significant. The assumptions for the test are fulfilled: the dependent variables (disfluencies) are continuous, the observations are independent (an interpreter cannot be represented both in the male and the female group) and the distributions of both groups (males and females) for each type of disfluency have a similar shape. The effect size (or r) for the Mann-Whitney's U test is calculated by dividing the z-value by the square root of n (the size of the sample).

2.3 Predictors

Based on the overview of the literature, pilot studies and metadata available in the corpus, several predictors besides the interpreter's sex have been identified as potentially influencing cognitive load and disfluencies and have therefore been chosen as predictors for the present study. Two predictors are categorical and were manually added to each transcription: source language and target language (either English, French or Dutch). The other predictors are continuous and were measured thanks to a tailor-made script after having been manually identified in the transcription. The Ear-Voice Span was measured by manually applying time tags to equivalent words uttered by the speaker and the interpreter and by calculating the duration in centiseconds between the two time tags. The source speaker's and interpreter's delivery rate were measured as the number of words uttered per normalized segment, excluding all types of disfluencies. The source speaker/interpreter speaking time ratio is the total speaking time of the source speaker (silent pauses excluded) divided by the interpreter's total speaking time (silent pauses excluded) and gives an impression of how much the interpreter actually speaks compared to the source speaker. Finally, all types of disfluencies were also measured for the source speaker in the same way as for the interpreter and their influence on the interpreter's disfluencies were measured. Moreover, the influence of the other disfluencies produced by the interpreter was analysed for each type of disfluency (e.g. the influence of the number of false starts produced by the interpreter on the number of filled pauses uttered by the interpreter). In total for each type of disfluency the effect of 16 predictors was analysed: interpreter's sex, source and target languages, Ear-Voice Span, source speaker/interpreter speaking time ratio, source speaker's and interpreter's delivery rate, number of filled pauses, lengthenings, false starts, silent pauses and the total duration of silent pause.

Mann-Whitney U tests are performed on each type of disfluencies to identify potential sex differences. In order to assess the influence of the 16 predictors, generalized linear mixed models were conducted for each type of disfluency. Mixed models allow for each interpretation to be included as a random variable in order to control for idiosyncratic effects. Contrary to disfluencies, which are measured both as averages per minute and per normalized segment, predictors are analysed in one condition only: frequency per normalized segment. Per-segment measurements are more accurate than averages measured for a whole speech. Given that most predictors are measurement per 10 seconds, the generalized linear mixed model was performed on the 10-second segments for disfluencies. Moreover, the authors want to assess the influence of predictors on the actual number of disfluencies instead of an average number of disfluencies. Since there are numerous zero values for each measurement of articulated disfluencies per 10 seconds, the count data Poisson regression was used in order to assess the influence of predictors on each type of disfluency. This method is indeed more robust with data containing numerous zero values. However, for the total duration of silent pauses, the data are continuous and a linear regression was used instead of a Poisson regression.

The results are presented through the β-coefficient, which indicates the individual contribution of each predictor to the regression model. The β-coefficient also indicates to what degree each predictor affects the outcome if the effects of all other predictors are held constant. They are interpreted differently between categorical and continuous variables. For continuous variables, each time the predictor increases by 1 unit, the disfluency increases or decreases by the number of units indicated by the beta value. For categorical predictors, one of the variables is taken as the reference variable and the others are compared to it. In order to estimate the goodness of fit of the model, the marginal R^2 and the conditional R^2 are also mentioned. The marginal R2 is the variance explained by the fixed effects, while the conditional R2 is the variance explained by both the fixed and the random effects.

3. Results

3.1 Descriptive statistics

As mentioned in Section 2, disfluencies were measured in two conditions: 1) normalized frequency per minute for each text and 2) normalized frequency per standardized segment of 10 seconds. Table 2 shows the number of disfluencies and the delivery rate in two conditions (normalized per minute and normalized per 10 seconds) for source speakers and interpreters.

Table 2. Disfluencies and delivery rate for source speakers and interpreters[4]

Variables		Source speakers				Interpreters			
		Mean	% of uttered tokens	M	SD	Mean	% of uttered tokens	M	SD
Number of words	per min	158.2	97.5%	156.9	19.4	142.1	91.2%	143.6	17.4
	per 10s	26.08	97.6%	25.76	5.14	23.63	92.1%	23.61	5.98
Number of filled pauses	per min	2.61	1.7%	1.18	3.56	7.52	5.3%	6.82	4.63
	per 10s	0.38	1.5 %	0.00	0.87	1.06	4.5%	0.98	1.23
Number of lengthening	per min	0.46	0.3%	0.00	0.88	3.36	2.4%	2.81	2.70
	per 10s	0.07	0.3%	0.00	0.33	0.54	2.3%	0.00	0.85
Number of false starts	per min	0.91	0.6%	0.85	0.70	1.50	1.1%	1.23	1.19
	per 10s	0.16	0.6%	0.00	0.41	0.25	1.1%	0.00	0.54
Number of silent pauses	per min	22.91		22.82	4.21	18.64		18.18	4.20
	per 10s	3.80		3.90	1.34	3.10		3.00	1.31
Average total duration of silent pauses (in sec)	per min	10.28		9.82	2.38	10.58		9.87	3.60
	per 10s	1.75		1.66	0.71	1.78		1.49	1.11

The most frequent type of disfluency in both interpreting and spontaneous speech is the silent pause followed by the filled pause. Interpreters clearly produce more disfluencies (filled pauses, lengthening and false starts) than source speakers. Counting only these three types of disfluencies, about 7.8% of uttered tokens are disfluent in interpreting,

---

[4] Sample size for per-min values is 180 and 4467 for per 10-sec values.

compared to about 2.6% for original speakers. The frequency of disfluencies in our corpus therefore appears to be slightly lower than that recorded in the experimental study by Christodoulides & Lenglet (2014). The average total duration of silent pauses is very similar between original speakers and interpreters, but the number of pauses is higher in spontaneous speech, which means that interpreters produce fewer but longer pauses than original speakers.

3.2 Sex differences in interpreters’ disfluencies

Table 3 shows the descriptive statistics for all types of disfluencies per interpretation and per sex and the results of the 10 one-tailed Mann-Whitney U tests performed on the 5 types of disfluency. The null hypothesis is that the distribution of the disfluency in both conditions will be identical. The significance level for p-values is 0.05.

Table 3. Descriptive statistics and sex differences in disfluencies[5]

Disfluency	Sex	Mean	Median	SD	Mann-Whitney U
Number of filled pause	F	6.95	6.25	4.20	U 3548
	M	8.10	7.33	4.10	p .076 Effect size (r) 0.11
Number of lengthening	F	3.01	2.27	2.79	U 3225.5
	M	3.71	3.39	2.56	p .009 Effect size (r) 0.18
Number of false start	F	1.60	1.37	1.29	U 3619.5
	M	1.41	1.14	1.08	p .109 Effect size (r) 0.09
Number of silent pause	F	18.98	18.90	3.94	U 3482
	M	18.29	17.31	4.44	p .052 Effect size (r) 0.12
Average total length of silent pause (in sec)	F	9.68	9.47	2.73	U 2982.
	M	11.49	10.89	4.11	p .001 Effect size (r) 0.23

Significant p-values were found for lengthenings (p=.009) and the total duration of silent pauses (p=.001), with higher values for men but small effect sizes (respectively 0.18 and

---

[5] N was180:(90 for females and 90 for males) .

0.23). No differences were found for the remaining disfluencies: number of filled pauses (p=.076), false starts (p=.109) and silent pauses (p=.052).

3.3 Influence of predictors

Generalized linear mixed models were conducted on each type of disfluencies with all predictors and the random effect “interpretation”. On top of the full model, the influence of predictors was also measured in single models between each predictor and the EVS.

In order to assess the actual influence of the predictors on the production of disfluencies at a given moment, the data used for the regressions are the 10-second measurements. Given the requirements for the Poisson regression, the data used as dependent variables (i.e. the disfluencies) are the count data (the raw frequencies of disfluencies per normalized segment). The sample size is therefore different from previous analysis, as some short and long segments were removed from the dataset. The predictors are the normalized values, as no differences were found when using the count data for the predictors. The predictors included in the regression are the following: interpreter’s sex, source and target languages, Ear-Voice Span, source speaker/interpreter speaking time ratio, source speaker’s and interpreter’s delivery rate, number of filled pauses, lengthenings, false starts, silent pauses and the total duration of silent pauses.

3.3.1 Interpreter's filled pauses

The sixteen predictors were included in the generalized linear mixed model. Results are presented in Table 4.

Table 4. Parameter Estimates for interpreter's filled pauses per 10 seconds

Predictors	β	Std. Error	z value	Sig.
Significant predictors				
Interpreter's sex: male	0.097	0.046	2.091	.037
Interpreter's sex: female	0			
Interpreter's language: French	0.223	0.057	3.919	<.001
Interpreter's language: English	0			
EVS	0.001	0.000	7.417	<.001
Interpreter's delivery rate	-0.057	0.004	-15.657	<.001
Interpreter's false starts	0.108	0.026	4.154	<.001
Interpreter's lengthenings	0.165	0.017	9.740	<.001
Interpreter's number of silent pauses	0.059	0.014	4.364	<.001
Interpreter's duration of silent pauses	-0.229	0.038	-6.047	<.001
Source speaker/interpreter speaking time ratio	-0.405	0.194	-2.083	.037
Source speaker's delivery rate	0.022	0.004	5.608	<.001
Source speaker's number of silent pauses	-0.041	0.015	-2.730	.006
Not significant predictors				
Interpreter's language: Dutch	0.091	0.056	1.626	.104
Source speaker's language: Dutch	0.076	0.067	1.130	.258
Source speaker's language: French	-0.062	0.072	-0.860	.390
Source speaker's false starts	-0.016	0.037	-0.450	.653
Source speaker's duration of silent pauses	-0.229	0.038	-6.047	.997
Source speaker's filled pauses	0.039	0.020	1.953	.051
Source speaker's lengthenings	0.021	0.051	0.409	.682

The marginal R^2 is 0.159 and the conditional R^2 is 0.304. One predictor did yield significant p-value in the single model, but do not have significant p-value in the full model: duration of silent pauses with p=0.997 (p=.001 for the single model). Five predictors do not have significant p-values in both models: interpreter's language: Dutch (p=.104), source speaker's language (p=.390 and p=.258), source speaker's false starts (p=0.653), lengthenings (p=.682), filled pauses (p=.051).

Three predictors did not yield significant p-values when included individually in the model but are significant in the full model: interpreter's sex with p=.037 (p=.265 for the single model), interpreter's number of silent pauses with p<.001 (p=.803 for the single model) and source speaker's delivery rate with p<.001 (p=.073 for the single model). Seven predictors (and partially the interpreter's language) have significant p-values in both models: interpreter's language French (p<.001), EVS (p<.001), interpreter's delivery rate p<.001), false starts (p<.001), lengthenings (p<.001), duration of silent pauses (p<.001), source speaker/interpreter speaking time ratio (p=.037) and source speaker's number of silent pauses (p=.006).

Female interpreters produce fewer filled pauses than male interpreters. French interpreters produce more filled pauses than English interpreters, while Dutch as a target language is not significant. Five predictors are associated with increased production of filled pauses by the interpreter: the EVS, the interpreter's number of false starts, lengthenings and silent pauses and the source speaker's delivery rate. Four predictors are negatively associated with the occurrence of filled pauses: the interpreter's delivery rate and duration of silent pauses, the source speaker's number of silent pauses and the source speaker/interpreter speaking time ratio.

3.3.2 Interpreter's lengthenings

Sixteen predictors were included in the generalized linear mixed model. Results are presented in Table 5.

Table 5. Parameter Estimates for interpreter's lengthenings per 10 seconds

	β	Std. Error	z-value	Sig.
Significant predictors				
Interpreter's sex: male	0.257	0.058	4.415	<.001
Interpreter's sex: female	0			
Interpreter's language: French	0.281	0.073	3.870	<.001
Interpreter's language: English	0			
Source speaker's language: French	-0.505	0.091	-5.535	<.001
Source speaker's language: English	0			
EVS	0.001	0.000	4.167	<.001
Interpreter's delivery rate	-0.087	0.005	-17.095	<.001
Interpreter's filled pauses	0.172	0.017	10.353	<.001
Interpreter's number of silent pauses	-0.047	0.019	-2.505	.012
Interpreter's duration of silent pauses	-0.230	0.046	-5.007	<.001
Source speaker's delivery rate	0.016	0.006	2.974	.003
Source speaker's filled pauses	0.097	0.026	3.794	<.001
Not significant predictors				
Interpreter's language: Dutch	-0.115	0.079	-1.460	.144
Source speaker's language: Dutch	-0.103	0.078	-1.317	.188
Interpreter's false starts	0.052	0.038	1.362	.173
Source speaker/interpreter speaking time ratio	-0.216	0.217	-0.998	.318
Source speaker's false starts	-0.015	0.051	-0.290	.772
Source speaker's lengthenings	0.078	0.066	1.181	.238
Source speaker's number of silent pauses	-0.020	0.020	-1.007	.314
Source speaker's duration of silent pauses	0.009	0.009	1.027	.305

The marginal R^2 is 0.186, and the conditional R^2 is 0.250. Three predictors did yield significant p-values when included individually in the model but are not significant in the full model: interpreter's false starts with p=.173 (p=.001 for the single model), source speaker/interpreter speaking time ratio with p=.318 (p=.007 for the single model) and source speaker's number of silent pauses with p=.314 (p=.037 for the single model). Five predictors do not have significant p-values in both models: source speaker's duration of silent pauses (p=.305), interpreter's language Dutch (p=.144), source speaker's language Dutch (p=.188), source speaker's number of false starts (p=.772) and lengthenings (p=.238).

Eight predictors (as well as the interpreter's and source speaker's languages to some extent) have significant p-values in both models: interpreter's sex (p<.001), interpreter's language French (p<.001), original speaker's language French (p<.001), EVS (p<.001), interpreter's delivery rate (p<.001), number of filled pauses (p<.001), silent pauses (p=.012) and average duration of silent pauses (p<.001), original speaker's delivery rate (p=.003) and number of filled pauses (p<.001).

Female interpreters produce fewer lengthenings than male interpreters. French interpreters produce more lengthenings than English interpreters, but Dutch as a target language is not significant. English as a source language triggers more lengthenings than French, Dutch is not significant. Four predictors are associated with increased production of lengthenings by the interpreter: EVS, interpreter's number of filled pauses, source speaker's number of filled pauses and delivery rate. The remaining three predictors are negatively associated with the production of lengthenings: interpreter's delivery rate, number and duration of silent pauses.

3.3.3 Interpreter's false starts

Sixteen predictors were included in the generalized linear mixed model. Results are presented in Table 6.

Table 6. Parameter Estimates for interpreter's false starts per 10 seconds

Predictors	β	Std. Error	z value	Sig.
Significant predictors				
Interpreter's language: French	-0.754	0.105	-7.186	<.001
Interpreter's language: Dutch	-0.871	0.099	-8.781	<.001
Interpreter's language: English	0			
Source speaker's language: Dutch	-0.318	0.112	-2.838	.005
Source speaker's language: English	0	.	.	.
EVS	0.001	0.000	2.738	.006
Interpreter's delivery rate	-0.032	0.008	-4.353	<.001
interpreter's filled pauses	0.116	0.026	4.411	<.001
Interpreter's duration of silent pauses	-0.236	0.070	-3.354	<.001
Source speaker's duration of silent pauses	-0.030	0.013	-2.337	.019
Not significant predictors				
Interpreter's lengthenings	0.047	0.039	1.207	.227
Source speaker's lengthening	0.010	0.111	0.090	.928
Interpreter's sex	-0.137	0.077	-1.784	.075
Source speaker's language: French	-0.195	0.118	-1.653	.100
Interpreter's number of silent pauses	0.017	0.028	0.615	.540
Source speaker's delivery rate	-0.006	0.008	-0.734	.463
Source speaker's filled pauses	-0.084	0.044	-1.910	.056
Source speaker's false starts	-0.002	0.075	-0.028	.978
Source speaker's number of silent pauses	0.045	0.029	1.527	.127
Source speaker/interpreter speaking time ratio	-0.111	0.324	-0.342	.734

The marginal R^2 is 0.055 and the conditional R^2 is 0.086. Two predictors did yield significant p-values when included individually in the model but are not significant in the full model: interpreter's lengthenings with p=.227 (p<.001 for the single model) and source speaker/interpreter speaking time ratio with p=.734 (p=.001 for the single model). Eight predictors do not have significant p-values in both models: interpreter's sex (p=.075), source speaker's language French (p=.100), interpreter's number of silent pauses (p=.540), source speaker's delivery rate (p=.463), source speaker's filled pauses (p=.056), false starts (p=.978), silent pauses (p=.127) and lengthenings (p=.928).

One predictor is not significant in the single model but is significant in the full model: source speaker's language Dutch with p=.005 (p=.081 for the single model). Five Predictors have significant p-values in both models: interpreter's language French (p<.001) and Dutch (p<.001), EVS (p=.006), interpreter's delivery rate (p<.001), filled pauses (p<.001), and duration of silent pauses (p<.001) and source speaker's duration of silent pauses (p=.019).

English interpreters seem to produce more false starts than Dutch and French interpreters, while English as a source language triggers more false starts than Dutch. Several predictors are positively associated with the production of false starts: EVS and interpreter's number of filled pauses. Other predictors have a negative influence on the production of false starts: the interpreter's delivery rate and duration of silent pauses, as well as the source speaker's duration of silent pauses.

3.3.4 Interpreter's silent pauses

As a reminder, two types of measurements were analysed for silent pauses: the frequency of silent pauses per normalized segment and the average total duration of silent pauses per normalized segment (i.e. not the average duration of a single silent pause, but the average total duration of all silent pauses per normalized segment). The same sixteen predictors were analysed for both types of measurement, but as the total duration of silent pauses is a continuous variable (and not a count data), a linear regression was used instead of a Poisson regression. The presentation of the results is therefore different.

The results for the frequency of silent pauses are presented in Table 7.

Table 7. Parameter Estimates for interpreter's number of silent pauses

Predictor	β	Std. Error	z value	Sig.
Significant predictors				
Interpreter's sex: male	-0.077	0.020	-3.787	<.001
Interpreter's sex: female	0			
Interpreter's language: French	-0.100	0.028	-3.565	<.001
Interpreter's language: English	0			
Interpreter's number of filled pauses	0.031	0.008	3.793	<.001
Interpreter's duration of silent pauses	0.274	0.020	13.761	<.001
Source speaker's number of filled pauses	-0.040	0.012	-3.410	<.001
Source speaker's number of silent pauses	0.044	0.008	5.667	<.001
Source speaker's duration of silent pauses	-0.026	0.004	-7.310	<.001
Source speaker/interpreter speaking time ratio	-0.742	0.098	-7.601	<.001
Not significant predictors				
Source speaker's number of lengthening	-0.026	0.029	-0.878	.380
Interpreter's language: Dutch	-0.045	0.026	-1.731	.083
Source speaker's language: French	0.053	0.029	1.805	.071
Source speaker's language: Dutch	0.024	0.029	0.826	.409
Interpreter's delivery rate	0.001	0.002	-0.010	.992
Interpreter's lengthenings	-0.018	0.012	-1.508	.132
Interpreter's false starts	0.007	0.016	0.446	.655
Source speaker's delivery rate	0.001	0.002	0.493	.622
Source speaker's false starts	0.010	0.021	0.475	.635
EVS	-0.000	0.000	-1.736	.083

The marginal R^2 is 0.116 and the conditional R^2 is 0.135. Three predictors did yield significant p-values when included individually in the model but are not significant in the full model: source speaker's language French with p=.071 (p<.001 for the single model), interpreter's delivery rate with p=.992 (p<.001 for the single model) and lengthenings with p=.132 (p=.005 for the single model). Six predictors predictors do not have significant p-values in both models: interpreter's language Dutch (p=.083), source speaker's language Dutch (p=.409), interpreter's false starts (p=.655), source speaker's delivery rate (p=.622), source speaker's false starts (p=.635), EVS (p=.083) and source speaker's lengthenings (p=.380).

Two predictors did not yield significant p-values when included individually in the model but are significant in the full model: interpreter's number of filled pauses with p p<.001 (p=.511 for the single model) and source speaker's filled pauses with p<.001 (p=.135 for the single model). Five predictors (and the interpreter's language to some extent) have significant p-values in both models: interpreter's sex (p<.001), interpreter's language French (p<.001), interpreter's duration of silent pauses (p<.001), source speaker/interpreter speaking time ratio (p<.001), source speaker's number of silent pauses (p<.001) and average duration (p<.001).

Female interpreters seem to produce more silent pauses than male interpreters. French interpreters produce fewer silent pauses than English interpreters. Three predictors are associated with increased production of silent pauses by the interpreter: interpreter's number of filled pauses and duration of silent pauses and source speaker's number of silent pauses. The remaining three predictors are negatively associated with the production of silent pauses: source speaker's number of filled pauses and duration of silent pauses and the source speaker/interpreter speaking time ratio.

The same predictors were analysed for the average total duration of silent pauses and results are found in Table 8.

Table 8. Effect of predictors on interpreter's duration of silent pauses

Predictor	β	Std. Error	df	t value	Sig.
Interpreter's sex: male	0.193	0.024	1182.528	8.1322	<.001
Interpreter's sex: female	0				
Interpreter's language: French	-0.079	0.031	2265.815	-2.558	.011
Interpreter's language: Dutch	-0.070	0.028	3455.451	-2.512	.012
Interpreter's language: English	0				
Source speaker's language: French	-0.111	0.038	519.856	-2.936	.004
Source speaker's language: Dutch	-0.101	0.035	864.554	-2.909	.004
Source speaker's language: English	0				
Interpreter's delivery rate	-0.058	0.002	4351.998	-31.041	<.001
Interpreter's filled pauses	-0.093	0.008	4332.238	-11.659	<.001
Interpreter's lengthenings	-0.094	0.011	4358.546	-8.260	<.001
Interpreter's false starts	-0.067	0.016	4333.803	-4.214	<.001
Interpreter's number of silent pauses	0.153	0.007	4357.972	21.684	<.001
Source speaker's delivery rate	0.014	0.002	4149.168	6.305	<.001
Source speaker's false starts	0.053	0.021	4325.042	2.576	.01
Source speaker's filled pauses	0.075	0.012	4049.241	6.369	<.001
Source speaker's lengthenings	0.123	0.029	4360.578	4.306	<.001
Source speaker's number of silent pauses	-0.064	0.008	4084.211	-8.046	<.001
Source speaker's duration of silent pauses	0.083	0.003	4120.363	29.162	<.001
Source speaker/interpreter speaking time ratio	2.249	0.037	4311.460	60.899	<.001
EVS	0.000	0.000	4271.201	2.735	.006

Three predictors did not yield significant p-values when included individually in the model but are significant in the full model: interpreter's language Dutch with p=.012 (p=.735 for the single model), source speaker's language Dutch with p=.004 (p=.466 for the single model) and source speaker's number of silent pauses with p<.001 (p=.529 for the single model). Thirteen predictors (and the interpreter's and source speaker's language to some extent) have significant p-values in the full model:

Male interpreters produce longer silent pauses than female interpreters. English interpreters produce longer silent pauses, while English as a source language triggers longer silent pauses. Eight predictors are associated with increased production of lengthening by the interpreter: the EVS, the interpreter's number of silent pauses, the source speaker's duration of silent pauses, delivery rate, frequency of false starts, filled pauses and lengthenings as well as the source speaker/interpreter speaking time ratio. The remaining five predictors are negatively associated with the production of lengthening: Interpreter's delivery rate, frequency of filled pauses, lengthenings and false starts and the source speaker's number of silent pauses.

4. Discussion

4.1 Sex differences

This study confirms results found in the literature: interpreters produce more filled pauses (Plevoets & Defrancq 2016), lengthenings (Tissi 2000), false starts (Pöchhacker 1995; Tissi 2000) and longer silent pauses (Tissi 2000; Cecot 2001; Ahrens 2005; Christodoulides 2013) than source speakers. Counting only these three types of disfluencies, about 7.8% of uttered tokens are disfluent in interpreting, compared to about 2.6% for source speeches. The average total duration of silent pauses per min is very similar for source speakers and interpreters, but source speakers produce a higher number of silent pauses than interpreters, which means that their silent pauses are shorter than interpreters' silent pauses.

The literature showed that men produce more fillers and repeats than women (Shriberg 1996; Bortfeld 2001). However Cecot (2001) found that female interpreters use more filled pauses and men use more and longer unfilled pauses. The results of the Mann-Whitney U test performed on our data give mitigated results. Male interpreters produce more lengthenings and have a higher total duration of silent pauses but no differences are found for the other types of disfluencies.

The mixed models confirm the results of the Mann-Whitney U test as they show that male interpreters produce more lengthenings and longer duration of silent pauses. However, the regressions also show that male interpreters produce more filled pauses and fewer silent pauses. As a reminder, the mixed models were performed on the per 10 second measurements while the Mann-Whitney U test was conducted per interpretation. The difference observed might therefore be due to the difference in measurement and to the fact that the influence of other predictors is included in the regression, while it is not in the Mann-Whitney U tests.

4.2 Influence of predictors

A summary of the influence of the 16 predictors on all disfluencies can be found in Table 9.

Table 9: Summary of the influence of predictors on disfluencies (beta values)

	Filled pause	Lengthening	False start	Silent pause	Duration of silent pause
Interpreter’s sex: male	0.097	0.257		-0.077	0.193
Interpreter’s sex: female	0	0		0	0
Interpreter’s language: Dutch			-0.871		-0.070
Interpreter’s language: French	0.223	0.281	-0.754	-0.100	-0.079
Interpreter’s language: English	0	0	0	0	0
Source speaker’s language: Dutch			-0.318		-0.101
Source speaker’s language: French		-0.505			-0.111
Source speaker’s language: English		0	0		0
EVS	0.001	0.001	0.001		0.001
Interpreter’s delivery rate	-0.057	-0.087	-0.032		-0.058
Interpreter’s filled pauses		0.172	0.116	0.031	-0.093
Interpreter’s lengthenings	0.165				-0.094
Interpreter’s false starts	0.108				-0.067
Interpreter’s number of silent pauses	0.059	-0.047			0.153
Interpreter’s duration of silent pauses	-0.229	-0.230	-0.236	0.274	
Source speaker/interpreter speaking time ratio	-0.405			-0.742	2.249
Source speaker’s delivery rate	0.022	0.016			0.014
Source speaker’s filled pauses		0.097		-0.040	0.075
Source speaker’s lengthenings					0.123
Source speaker’s false starts					0.053
Source speaker’s number of silent pauses	-0.041			0.044	-0.064
Source speaker’s duration of silent pauses			-0.030	-0.026	0.083

Female interpreters produce fewer filled pauses and lengthenings and have a shorter average duration of silent pauses, but produce more silent pauses than male interpreters. There is no sex difference for false starts. French and English as target languages are statistically significant for all types of disfluencies.

While French interpreters produce more filled pauses and lengthenings than English interpreters, they produce fewer false starts, silent pauses and have a shorter duration of silent pauses. Dutch interpreters produce fewer false starts and have shorter duration of silent pauses than English interpreters, but no significant differences are found for the other types of disfluencies. When it comes to the influence of source languages, French and Dutch trigger shorter duration of silent pauses than English, Dutch triggers fewer false starts and French triggers fewer lengthenings than English. The fact that Dutch as a source language is not significant for most disfluencies and triggers fewer false starts and shorter duration of silent pauses contradicts the literature according to which language pairs with an SOV constituent order in the source language (such as Dutch) increase cognitive load (Seeber 2011). French as a source language seems therefore not to trigger many disfluencies, while English does.

The Ear-Voice Span is mentioned n the literature as a predictor of disfluencies (Tóth 2011) and its significance is confirmed for four types of disfluencies: filled pauses, lengthenings, false starts and the duration of silent pauses. The positive effect of EVS can have two potential explanations: 1) the more the interpreter hesitates, the longer it takes to produce a sentence, the longer the EVS become; 2) a longer EVS is often associated with high cognitive load; similarly disfluencies are also associated with high cognitive load. Their positive correlation could therefore be interpreted in terms of both variables' individual association with cognitive load. The interpreter's delivery rate is negatively associated with all disfluencies expect for the frequency of silent pauses. This seems logical as disfluencies will tend to decrease the fluency, and therefore the delivery rate, of the interpreter (as a reminder, articulated disfluencies are not included in the word count for delivery rate). The source speaker/interpreter speaking time ratio has a negative influence on two types of disfluencies (number of filled and silent pauses) but a positive influence on the duration of silent pauses. This means that when the source speaker speaks more than the interpreter, the interpreter produces fewer filled pauses and silent pauses, but has a longer duration of silent pauses. Less speaking time for the interpreter logically means less opportunity for disfluencies. Similarly, if the duration of silent pauses increases, it means that the speaking time decreases, since silent pauses are not included in the speaking time.

The production of filled pauses is increased by the production of lengthenings, false starts and silent pauses by the interpreter, and also triggers an increase in the production of the other types of disfluencies (except for the duration of silent pauses). However the production of lengthenings and false starts do not mutually trigger an increase. The production of lengthenings is decreased by silent pauses while they have no influence on false starts. The interpreter's duration of silent pauses is negatively influenced by all types of uttered disfluencies and vice versa. This might simply be explained by the fact that longer duration of silent pauses decrease the time dedicated to speech, and therefore also reduces the number of articulated disfluencies. Similarly, the production of silent pauses logically increases the duration of silent pauses.

The source speaker's delivery rate was mentioned in the literature (Plevoets & Defrancq 2016; Seeber 2011) as increasing the cognitive load and the number of filled pauses. Our data show that this predictor indeed has a positive influence on three types of disfluencies: filled pauses, lengthenings and the duration of silent pauses. The source speaker's uttered disfluencies have a positive influence on the duration of silent pauses by the interpreter. The source speaker's number of filled pauses has a positive influence on the interpreter's lengthening, but a negative influence on the production of silent pauses. The source speaker's number of lengthening and false starts only have a positive influence on the duration of silent pauses. The source speaker's number of silent pauses decreases the production of filled pauses and the duration of silent pauses, but increases the frequency of silent pauses.

5. Conclusion

The aim of this study was to identify the variables determining the frequency of four types of disfluencies in interpretations: lengthenings, filled pauses, false starts and silent pauses.We particularly focused on the interpreter's sex as a predictor, because the literature indicates that women and men perform differently on linguistic tasks. In many of the studies, women are found to perform better than men. Meta-studies, on the other hand, generally fail to confirm these observations. Several other predictors known to increase cognitive load in interpreters and, therefore, likely to trigger disfluencies in interpreters, were also examined: these are variables relating to language pairs (source language and target language), properties of the source and the target texts (delivery rate) and their relation in time (EVS) and, finally, to the occurrence of other types of disfluencies both in speakers and in interpreters).

Whether female interpreters produce fewer disfluencies because their higher verbal skills might mean that they are less subjected to cognitive load. The results show that interpreters indeed produce more filled pauses, lengthenings, false starts and longer silent pauses than source speakers. This confirms the general idea that disfluencies are caused by cognitive load: interpreting is assumed to be cognitively more demanding than spontaneous speech and is therefore expected to be more prone to disfluencies. When it comes to sex differences and women's greater fluency, the results also tend to confirm our tentative hypothesis.

Both the Mann-Whitney U test and the mixed models show that the hypothesis is confirmed for lengthening and the duration of silent pauses but that no difference exist for false starts. Results are more mitigated for filled pauses and silent pauses but given the increased accuracy of the mixed models, both in terms of measurement and inclusion of other predictors, we can consider that its results are more reliable and that men also produce more filled pauses, but fewer silent pauses.

It is important to remember that the higher frequency of disfluencies in male interpreters is not necessarily a bad sign. As mentioned in the introduction, disfluencies are not viewed as errors but as solutions to errors in speech planning (Betz et al. 2015) and they can play a positive role by helping the identification of upcoming words for example (Howell & Young 1991; Fox Tree & Schrock 1999; Brennan et al. 2001; Fox Tree 2001; Arnold et al. 2003; Arnold et al. 2004; Ferreira et al. 2004). It could, however, be useful for trainers to pay particular attention to the types of disfluencies produced by students and determine what the causes are and if a particular attention is needed to fix them.

As for the textual and contextual variables explored in our study, as potential triggers of cognitive load and therefore of disfluencies, our results show that they do not always have an influence on the production of disfluences.

The following predictors tend to increase the production of most disfluencies: the Ear-Voice Span, source speaker's and interpreter's number of filled pauses and source speaker's delivery rate. Other predictors are negatively associated with most disfluencies: interpreter's delivery rate and duration of silent pauses. The other predictors yielded mixed results. The source speaker/interpreter speaking time ratio decreases the production of filled and silent pauses but increases the duration of silent pauses. Interpreter's lengthenings and false starts increase the production of filled pauses but decrease the duration of silent pauses. The interpreter's number of silent pauses increases filled pauses and the duration of silent pauses, but decreases lengthenings. Source speaker's lengthenings and false starts only increase the duration of silent pauses, while the Source speaker's number of silent pauses decreases the number of filled pauses and the duration of silent pauses.

Dutch as a target or source language is only significant in some cases, and in non-predictable ways: different constituent orders, as between Dutch and one of the other languages, do not seem to trigger disfluencies. English seems to trigger the most disfluencies and French the least. French interpreters produce more filled pauses and false starts but fewer lengthenings and silent pauses.

It is important to mention that some triggers of disfluencies have not been taken into account in this research project, because they are difficult to operationalize and require a depth of analysis that could not be attained in the framework of this research: new information (Clark & Fox Tree 2002; Arnold et al. 2003), heavy constituents (Swerts 1998; Arnold et al. 2000; Watanabe et al. 2008) or long sentences (Oviatt 1995; Shriberg 1996; Gile 2008; Chmiel & Mazur 2013), the position of disfluencies in sentences (Boomer 1965; Shriberg 1996), target lexical density (Plevoets & Defrancq 2016) and the topic or domain (Schachter et al. 1991). The results might therefore be different when these predictors are included. Moreover, the aim of this study was to compare several types of disfluencies between male and female interpreters, and not to give a comprehensive overview of the types of disfluencies in interpreting. As a consequence, not all types of disfluencies were taken into account, such as for instance repetitions, grammatical errors, articulation rate and alterations in voice quality (Jameson et al. 2009; Yap 2012; Schuller & Batliner 2013). Obviously, the inclusion of more variables would greatly enhance our understanding of disfluencies and the associated cognitive load in interpreters. We would also wholeheartedly welcome complementary experimental research on disfluencies in interpreting to confirm or contradict our corpus-based findings in highly controlled circumstances.

Chapter 4: Interpreting into an SOV Language: Memory and the position of the Verb. A Corpus-based Comparative Study of Interpreted and Non-mediated Speech

Abstract

In Dutch and German subordinate clauses, the verb is generally placed after the clausal constituents (Subject-Object-Verb structure) thereby creating a middle field (or verbal brace). This makes interpreting from SOV into SVO languages particularly challenging as it requires further processing and feats of memory. It often requires interpreters to use specific strategies (e.g. anticipation) (Lederer 1981; Liontou 2011). However, few studies have tackled this issue from the point of view of interpreting into SOV languages. Producing SOV structures requires some specific cognitive effort as, for instance, subject properties need to be kept in mind in order to ensure the correct subject-verb agreement across a span of 10 or 20 words. Speakers therefore often opt for a strategy called extraposition, placing specific elements after the verb in order to shorten the brace (Hawkins 1994; Bevilacqua 2009). Dutch speakers use this strategy more often than German speakers (Haeseryn 1990). Given the additional cognitive load generated by the interpreting process (Gile 1999), it may be assumed that interpreters will shorten the verbal brace to a larger extent than original speakers.

The present study is based on a corpus of interpreted and non-mediated speeches at the European Parliament and compares middle field lengths as well as extraposition in Dutch and German subordinate clauses. Results from 3461 subordinate clauses confirm that interpreters of both languages shorten the middle field more than original speakers. The study also shows that original German speakers have longer middle fields than Dutch original speakers, but this is not the case for interpreters. Dutch and German interpreters appear to use extraposition partly because they imitate the clause word order of the source speech, showing that, in this case, extraposition can be considered an effort-saving tool.

Keywords : Corpus-based interpreting studies, cognitive effort, SOV, verbal brace, extraposition

1. Introduction

Simultaneous interpreting is a demanding cognitive task (Gile 1999; Seeber 2011) in which interpreters need to find a balance between several simultaneous tasks: listening, analysis, memory, production and coordination. The different components of memory (short term, long-term and working memory) play a key role in this process (Moser 1978; Darò and Fabbro 1994). Interpreters therefore tend to develop strategies to reduce their cognitive effort.

One of the most frequently mentioned contexts in which interpreters are said to need interpreting strategies is interpreting between languages with different word orders, in particular SOV-SVO. In one of the earliest studies on ear-voice-span (EVS), Oléron and Nanpon (1965/2002) find that EVS is shorter when interpreters work from English into French (both SVO) than when they work from German (SOV) into French (SVO). Goldman-Eisler (1972) attributes this difference to the position of the verb in German. Lederer (1981) refers in particular to word order differences when discussing *anticipation* as a strategy: as SVO languages need the verb to occur earlier in the sentence than SOV languages, interpreting from an SOV language (German) into an SVO language (French) puts the interpreter before a crucial strategic dilemma: either wait for the verb and halt delivery of all intermediary information, or try to guess what verb will be used and anticipate it. In his study on SOV-SVO interpreting, Bevilacqua (2009) offers a more comprehensive overview of strategies employed by interpreters to overcome structural differences between languages. Analysing interpreters' output, he finds evidence of anticipation, ear-voice span management, reformulation, compression and omissions. In terms of cognitive load attached to different strategies, Seeber (2011) models four different possible interpreter responses to SOV input while interpreting into an SVO language: waiting, stalling, chunking and anticipating. The first three strategies come with a high memory load, while the last one comes with an inherent risk of errors and subsequent cognitive load at the output end, if errors have to be corrected.

Studies in interpreting research have mainly tackled SOV-related strategies from the perspective of the recipient, focusing on the question how interpreters *respond* to SOV. With the exception of EVS-related issues, such as discussed in Lee (2002), the issue of how interpreters *produce* SOV structures on the basis of an SVO input does not seem to have drawn interest from the research community. Yet, it may be expected that word order differences pose particular challenges to interpreters in both directions. Interpreters producing an SOV structure from an SVO input need to delay delivery of the verb until they reach the clause's particular final slot. This requires memory capacity and elicits cognitive load.

The lack of interest in SOV output in interpreting studies mirrors a more general tendency in the literature on memory: Acheson and McDonald (2009), for instance, express surprise over the one-sided focus on acquisition and reception in studies on working memory, as it is generally known that language production requires particular demands in terms of verbal working memory. However, early research on the effects of increased memory load and memory impairment on the production of spontaneous speech failed to produce significant results (Sternberg 1969; Shallice and Butterworth 1977; Klapp, Greim, et al. 1981; Vallar and Baddeley 1984), which may have discouraged most of the research community from pursuing research in that direction.

In this paper, we set out to investigate interpreters' production of particular SOV patterns, i.e. verbal braces in subordinate clauses, to find out if the cognitive load inherent in the interpreting process affects the interpreters' output. We will focus particularly on the point of delivery of the verb within the pattern, comparing non-mediated language production (i.e. original speeches delivered at the European Parliament) and interpreted language. In section 2, we first describe the features of the SOV patterns in the languages under study, German and Dutch. Section 3 then combines this description with fundamental assumptions in interpreting to predict features of SOV patterns in interpreters' output. In section 4, we present the corpus materials used for this study and the methodology used to compare non-mediated output and interpreters' output. Results and conclusions drawn from the results are given in sections 5 and 6.

1.1. German and Dutch as SOV languages

Dutch and German are often referred to as SOV languages (Gerritsen and Stein 1992, Eisenberg 1994), because the verbal group can be placed at the end of a clause. In canonical sentences with a simple verb, verbs are normally in second position following the subject or any other phrase, as illustrated in example (1), where *vandaag* ('today') is a temporal adjunct.

1) Vandaag werkt hij in de stad
[today works he in town]

In canonical sentences with a complex verb and in subordinate clauses, the (main) verb is situated near the end of the sentence, as in example (2).

2) Es ist besonders wichtig, dass wir den Dialog mit Russland fortführen
[it is particularly important that we the dialog with Russia continue]
(EPICG_20080810_formal sitting1_Ingrid Betancourt_I_de)

This structure creates a *verbal brace or bracket construction* (Drach 1937/1963, Van Haeringen 1947, Haeseryn, Romijn, *et al.* 1997, Zwart 2011: 31-79) that divides the sentence into several fields. The verbal brace in example (2) starts with the subordinating conjunction dass, hereafter also called left pole, and ends with the verb *fortführen*, hereafter called right pole. Three fields can be distinguished: the pre-field (*Es ist besonders wichtig*), the middle field (*wir den Dialog mit Russland*) and the after field, which is empty in this case. In written German and Dutch subordinates, the unmarked pattern consists of a middle field for clausal constituents, followed by the verb in clause-final position and an empty after field (Thurmair 1991: 175ff). Although the middle field (i.e. the field found inside the verbal brace) is considered the standard slot for clausal constituents both in German and Dutch, some of these constituents can occur in the after field. Their position is governed by a complex interaction of pragmatic, thematic, lexical and syntactic parameters (e.g. Jansen 1978; Vandenbosch 1985; Koops 1986; Braecke 1990; Zwart 1990; de Schutter 2003; Coussé 2009; Willems 2017). Long and complex clausal constituents, for instance, tend to be placed in the after field position, at least partially (Haeseryn, Romijn, *et al.* 1997). This phenomenon is called extraposition and is illustrated in example (4), which is an adapted version of the real corpus example in (3):

3) onze fractie denkt dat de enige manier om te zorgen voor stabiliteit euh en zekerheid aan beide kanten van de Atlantische Oceaan en te zorgen voor een evenwichtige samenwerking tussen de Europese Unie en de Verenigde Staten deze weg is
[our group thinks that the only way to ensure stability and safety on both sides of the Atlantic Ocean and to ensure a balanced cooperation between the EU and the United States this way is]
(EPICG_20080810_formal sitting1_Ingrid Betancourt_I_nl)

4) onze fractie denkt dat deze weg de enige manier is om te zorgen voor stabiliteit euh en zekerheid aan beide kanten van de Atlantische Oceaan en te zorgen voor een evenwichtige samenwerking tussen de Europese Unie en de Verenigde Staten
[our group thinks that this way the only way is to ensure stability and safety on both sides of the Atlantic Ocean and to ensure a balanced cooperation between the EU and the United States]

While in example (3), the middle field contains 34 words and there is no after field, in example (4) the middle field is reduced to 5 words, and an after field of 29 words is created. In between these two extremes, other options are available: is could, for instance, also be inserted just before *en te zorgen*.

Extraposition is available for most syntactic constituents except subject and object, where the length and complexity of the constituents increases the likelihood of extraposition. However, not only the length of constituents plays a role; the length of the middle field and of the after field also determine constituent placement. The longer the middle field, the more likely it is for constituents to be extraposed. Willems (2017) shows that this “overflow” capacity of the after field is not unlimited either: the longer it becomes, the less likely it is to be expanded yet further. Other linguistic parameters that have been found to promote extraposition in spoken Dutch are grammatical function (prepositional phrases are more prone to extraposition than adverbial adjuncts: Jansen 1978; Braecke 1990), and the indefiniteness of the noun included in the prepositional phrase (Jansen 1978). Extra-linguistic parameters are formal register, higher social class, older age and male gender (Jansen 1978; Braecke 1990).

Although Dutch and German have in principle the same possibilities to extrapose constituents, Dutch language users seem to do it more systematically than German speakers, who tend to preserve verbal braces instead (Haeseryn 1990). German therefore has a more rigid verb-final structure, while Dutch can be described as a “moderately verb-final SOV language” (De Schutter 1994: 466). In the interpreting data we used for this study, braces such as in example (3) appeared to be more typical for German than for Dutch usage.

1.2 SOV and extraposition in interpreting

The challenges of interpreting from an SOV language into an SVO language are generally known and the effects on interpreter output are easily predictable: longer EVS and associated increased cognitive load, pausing, and anticipation. The question that interests us here is what to expect when interpreters are working from an SVO language into an SOV language.

It seems reasonable to assume that some additional memory effort will be needed, as interpreters will have to postpone delivery of the verb if they are interpreting from an SVO language. This is a cross-linguistic memory effort: its intensity partly depends on the position of the verb in the source language. However, not only the verb must be stored in memory, but also agreement features. In example (3), for instance, the verb is, which agrees in person and number with its subject *de enige manier*, is separated by 26 words from its subject. The agreement features must consequently be stored in memory over a span of 26 words. Storing agreement features is a monolingual memory effort: its intensity depends on the position of the subject in the target language. Agreement features have been shown to be vulnerable in SOV contexts, because of items occurring between the subject and the verb (Vigliocco 1998). The sooner a speaker delivers the verb, the lower the risk of agreement errors. The cross-linguistic effort is required of interpreters only; the monolingual memory effort concerns all users of the target language.

All in all, the total memory effort spent to interpret from an SVO language into an SOV language is likely to be smaller than when interpreting from SOV into SVO, as only the verb and agreement features have to be kept in memory and not all the clausal constituents that happen to occur in a verbal brace. In addition, extraposition offers interpreters a structural means of memory management: with some restrictions, constituents can be extraposed, narrowing down the gap between the verb in the target SOV language and its equivalent in the SVO source language, and between the verb and its subject in the target language. In other words, extraposition offers researchers a view on memory management in interpreters: comparing the point at which the verb is delivered in non-mediated language and in interpreting may reveal discrepancies that are significant for memory management in interpreters. As said, original speakers only face the memory effort related to the agreement features of the verb, while interpreters also have to cope with the cross-linguistic memory effort. To reduce the total memory effort, interpreters could be tempted to deliver the verb sooner than original speakers and, consequently, extrapose more often.

Before this hypothesis can be tested, we first need to check whether the parameters that determine extraposition in non-mediated language are likely to interact with interpreting processes in ways that would distort the results of a comparative analysis. As described above, the position of the verb in German and Dutch SOV structures is determined by a variety of linguistic and extra-linguistic factors. Among the linguistic factors, constituent length and complexity is likely to be different in non-mediated language and interpreting. Interpreting involves higher cognitive load, which interpreters are known to compensate by producing fewer long and complex constituents than non-interpreters, either because they omit non-essential source units (Shlesinger 2003) or because they apply some form of chunking (Meuleman and Van Besien 2009; Seeber 2011) to distribute the information over more but simpler constituents. As extraposition is more likely in the case of long and complex constituents, lower rates of long and complex constituents in interpreting could skew the data in favour of structures where the verb actually occurs later in interpreting than in non-mediated speech. Regarding the other linguistic factors, i.e. indefiniteness and grammatical function, there is no evidence to suggest that their frequencies are different in interpreted and non-mediated speech. They will not be considered as confounding variables in this study.

As far as the extra-linguistic factors are concerned, sex could be a distorting factor, as the interpreting profession is predominantly female. Women have been shown to be more reluctant to extrapose constituents than men (Jansen 1978). Interestingly, this can be related to memory management. The female advantage in memory tasks is widely documented: women generally perform better than men in verbal and nonverbal memory tasks (Kramer, Delis, *et al.* 1997; Loonstra, Tarlow, et al. 2001; Kimura and Seal 2003; Maitland, Herlitz, *et al.* 2004). Assuming memory is a factor in extraposition, the greater propensity of men to extrapose constituents could be due to memory capacity. The question whether this sex difference is likely to show in interpreting data is undecided. The little research that has been carried out on differences between male and female interpreting has focused on sociolinguistic rather than cognitive properties (Magnifico and Defrancq 2016; 2017, Russo 2018). Mason (2008) found evidence for sex differences in omission rates for consecutive interpreters in the courtroom and Baes (2012) and Defrancq (2013) discovered longer Ear-Voice-Span for female interpreters at the European Parliament. Cecot (2001) found that women used more filled pauses and men used more unfilled pauses. As far as memory differences are concerned, one could argue that these are likely to widen in activities such as interpreting, which put a particular strain on memory. On the other hand, it might also be argued that the intensive training which interpreters are offered, including memory training, is likely to close the gap between the sexes.

Besides sex, age and social class have also been shown to influence extraposition. Neither of these parameters is easy to control, for in interpreting research the number of participants in experimental research is always very limited, while in corpus research, data on the age and social class of interpreters is usually not available. We will assume that age and social class are not confounding variables in our study. The final parameter, formality, on the contrary, will have to be taken into account: interpreters usually operate in official contexts and are expected to use formal registers, which again could counterbalance a tendency to extrapose constituents.

Finally, in relation to interpreting in particular, it is also important to refer to source text influence as a factor determining the position of the verb. In a corpus-based study, De Sutter and Van de Velde (2008) found that Dutch translators of German prose extraposed significantly fewer prepositional phrases than Dutch authors. The opposite was however not true: German translators of Dutch prose did not extrapose more than German authors. This means that the source language can have an influence on the choice to extrapose constituents, but this influence could in turn be reinforced or inhibited by other factors. In our study, French is the source language for both the Dutch and German interpretations. French being an

SVO language, all major constituents follow the verb, which could be an important trigger for extraposition.

In sum, the main hypothesis of the present study is that extraposition is more frequent in interpreting than in non-mediated speech, because interpreters have an interest in delivering the verb sooner than original speakers. Some distorting factors that have been identified point towards less extraposition (a majority of female speakers and more formal registers), while the occurrence of specific interpreter strategies could lead to more extraposition (e.g. anticipation).

2. Data and methodology

2.1. Corpus

The data we use in this study are drawn from the European Parliament Interpreting Corpus Ghent (EPICG), supplemented with German data taken from the same context.
The choice of corpus data in interpreting research and, in particular corpus data from the European Parliament, is disputed. Diriker (2004), for instance, states that:

> The online availability of the speeches and their interpretations at the EP's plenary sessions is certainly an invaluable source for researchers interested in analyzing authentic corpora of interpreting. Caution, however, is necessary, since the online availability of such recordings means they can be used by everyone, including researchers who have never seen the European Parliament in session nor talked to the interlocutors there to gain an idea of the constraints of interpreting in that particular setting. Although analysis of any data will by nature never be a mirror reflection of reality, drawing conclusions on SCI [Simultaneous Conference Interpreting] as situated action based on de-contextualized recordings must be taken with an even larger grain of salt. (Diriker 2004:215)

Nevertheless, interpreting corpora are considered by many to have the potential to reinforce the empirical foundations of interpreting research (Shlesinger 1998). Corpus data are naturalistic data produced in a real-life environment by professionals and therefore reflect the interpreting activity in a way experimental data cannot. Corpora also give a more comprehensive view of the interpreting process and are more likely to produce reliable generalizations about the various interpreting activities.

The use of corpora is becoming increasingly popular in interpreting studies. New technologies have offered solutions to the time-consuming compilation, transcription and analysis of interpreting data. For a number of years, the plenary sessions and some of the committee sittings of the European Parliament can be downloaded from the website of the European Parliament. Plenary sessions may not be an entirely adequate response to the shortcomings of experimental data, as the interventions are generally very short (1 to 6 minutes), the source delivery rate is very high (155 words per minute on average), the speeches are often read out, and the working conditions only reflect this specific working environment. However, the amount of data allows for representative studies to be carried out on the context itself. The institutionalized setting of the debates furthermore ensures the homogeneity of the data in regard to the speakers' allocated time and the interpreters' qualification and working conditions. In addition, as both the non-mediated and the interpreted data are produced in the same context, it can reasonably be expected that there are no significant differences in levels of formality between both types of data. This ensures that one of the distorting factors mentioned in section 3 – i.e. formality – is controlled for. In the context of this study, we did not deem it necessary, as suggested by Diriker (2004), to re-contextualize the corpus data: the metadata collected as part of the corpus compilation effort

(Bernardini, Collard, *et al.* 2017) suffice to gain a good understanding of the working conditions of EP interpreters.

The corpus is still being compiled at the time of writing of this study. Speeches are randomly selected on the European Parliament's website from plenary sessions between 2008 and 2013 and transcribed according to the Valibel instructions (Bachy, Dister, et al. 2007). Besides the words spoken, other acoustic features are included in the transcriptions (e.g. disfluencies, self-repairs, false starts, filled and silent pauses). In its current shape, the corpus contains approximately 200,000 word tokens, with both source and target speech in English, French and Dutch. About half of the data of the *EPICG* are compiled in *EXMARaLDA* (Schmidt 2012). *EXMARaLDA* has the considerable advantage that transcriptions and audio can be aligned and that the software comes with a dedicated corpus compilation module and a concordancer.

The present study is based on a sub-corpus of *EPICG* and compares interpreted speeches with non-interpreted speeches. It comprises German and Dutch original speeches and German and Dutch interpretations of French source speeches. The Dutch data were extracted from the corpus; the German data were compiled specifically for this study. It was crucial to use interpretations from a third language, as using Dutch interpretations from German or vice versa could have yielded distorted data. De Sutter and Van de Velde (2008) indeed conclude on the basis of a literary corpus that Dutch translators appear to be influenced by the German source texts they translate, as verbal brace lengths are considerably longer in Dutch translations than in Dutch non-mediated literary production.

The sub-corpus contains 63 original speeches (39 speeches in German and 24 speeches in Dutch) and 80 interpretations from French (41 speeches in German and 39 speeches in Dutch). The total amount of tokens is 89,334 (26,450 tokens for German original speeches; 20,130 tokens for German interpretations from French; 19,200 tokens for Dutch original speeches and 23,554 tokens for Dutch interpretations from French). An attempt was made to create a balanced corpus by having the same number of mediated and non-mediated speeches, and by having the same French orginal speeches for both German and Dutch interpretations. However, it appears that the number of subordinates is higher in Dutch speeches. Therefore, the authors decided to add German mediated and non-mediated speeches in order to have a more balanced amount of subordinates in each group. The final corpus remains slightly unbalanced because it is almost impossible to predict how many subordinate clauses will be contained in a speech. Moreover, several speeches had to be deleted as it appeared that the interpreter was on relay, and not interpreting the speech directly from the source speech.

2.2. Data selection

2.2.1. General principles

The present study focuses on subordinate clauses because they always contain two poles, which is necessary to measure the length of the middle and after fields. In order to be able to clearly delineate the middle and after fields, only constructions starting with a subordinating conjunction (e.g.: *Ik zei dat Jan morgen op tv komt in zijn eentje.*) or relative clauses (e.g.: *Hast Du die Lampe, die du gestern gesehen hast, gekauft?*) have been taken into account. All relevant subordinates in the transcripts of German and Dutch originals as well as of interpreted speech were manually annotated by two independent researchers. Some subordinates were incomplete, because either the left pole or (parts) of the verbal group were missing. Thus in example (5), a subordinate is clearly initiated by the subordinating conjunction *dass*. However, instead of completing the first clause with a verbal group, another subordinate is added. Such clauses were excluded from the dataset used for this study.

5) …dass bei bestimmten wichtigen Punkten / die vor allem zu tun haben mit der Einrichtung einer europäischen Aufsichtsbehörde / oder auch mit / der / europäischen Registrierung und Zulassung euh euh von euh Wertpapierverwaltungsgesellschaften
[…that for certain important points / which mainly have to do with establishing a European supervisory body / or also with European registration and certification of portfolio management companies]
(EPICG_20082209_Hedgefunds private equity_Pierre Jonckheer_I_nl)

The dataset also included examples where two or more subordinates shared the same left pole. In example (6), for instance, the relative pronoun *der* triggers three subordinates:

6) …der das Thema mit angesprochen hat mit auf die Tagesordnung gesetzt hat und damit den Beitrag geleistet hat
[… who the topic mentioned on the agenda put and a contribution made]
(EPICG_ 13012015_ Review of the Italian Presidency (debate)_Manfred Weber_O_de)

Here, the number of words contained in the middle field was counted for each subordinate. We labelled this variable the “real middle field” (RMF). We analysed examples such as (6) as consisting of three different subordinates for which the RMF only contains the clausal constituents related to the verbal group. For this case: the RMF for the first subordinate is *das Thema mit*, for the second subordinate *mit auf die Tagesordnung*, and for the third *damit den Beitrag*. Items that could not have been used if there were only one subordinate clause, like the coordinating conjunction *und* in this example, were not included in the count.

After the filtering process, a total of 3461 instances of verbal braces in subordinate clauses were included in the study and distributed as shown in Table 1.Interpreters were found to produce more subordinate clauses than speakers. While subordinate clauses could be considered as more complex and therefore be avoided by interpreters, interpreters might simply be copying the subordinates in the French source speech or use very short and simple subordinate clauses.

Table 1. Distribution of subordinate clauses in the corpus

	German non-mediated	German interpreted	Dutch non-mediated	Dutch interpreted
No. of Token	26,450	20,130	19,200	23,544
No.of Subordinates	602	1015	659	1,185
Subordinates per 100 Token	2.28	5.04	3.43	5.03

The RMF was considered a proxy of the memory effort that was spent to produce the structure. Although using word counts to measure memory effort is not an ideal solution (in general, “chunks” or “information units” are considered better tools in this respect), these latter units lack precise definitions, and are not easy to identify in speech, where punctuation does not exist. Given this lack of clarity, for this study the word was chosen as a generally replicable unit.

However, it is also true that a more accurate representation of memory effort would be the span of words separating the subject from the verb. After all, the difference between SVO and SOV word order does not involve the position of the subject. Moreover, agreement features expressed on the verb are determined by the subject, and must be kept in memory as soon as the subject is articulated. We nevertheless opted for the RMF because quite a number of the examples extracted are subjectless, either because the clause is infinitival, or because its subject is the antecedent of a relative pronoun and thus outside the brace.

All fully pronounced words other than the right and left poles were counted. Hesitation markers such as *euh*, as in example (7), truncated words or clear self-repairs, as in (8), were not included in the word count.

7) …$\boxed{\text{dass}}\ \underbrace{\text{euh / Fremdsprachen}}_{\text{RMF = 1}}\ \boxed{\text{erlernt werden können}}$

[…$\boxed{\text{that}}$ euh / foreign languages $\boxed{\text{can be learnt}}$]

(EPICG_20080209_Explanations of vote_Astrid Lulling_I_nl)

8) …$\boxed{\text{dass}}\ \underbrace{\text{die eigenen Krea/ Kreaturen}}_{\text{RMF = 3}}\ \boxed{\text{entarten}}$

[…$\boxed{\text{that}}$ the own cre/ creatures $\boxed{\text{degenerate}}$]

(EPICG_20080810_formal sitting1_Ingrid Betancourt_I_nl)

RMF length provides us with an indication of memory effort, but it does not inform us on the use of strategies to manage memory capacity. As explained above, extraposition is viewed here as one of the strategies interpreters could use to avoid storage of information over comparatively long stretches of speech. We therefore counted the number of corpus examples in which extraposition is observed in non-mediated and in interpreted speech. For this, it was necessary to distinguish between items in the after field that theoretically could (or even should) be placed in the middle field and items that could not: only the former can be considered extraposed. We therefore employed the concept of the theoretical middle field (TMF) – comprising both the items found in the observed real middle field (RMF) and those in the after field (AF) that could have been placed in the middle field[1].

---

[1] Additional analyses were carried out with a stricter definition of the TMF (excluding all subordinate clauses from the TMF). Since the results were very similar, these analyses are not mentioned here.

Example (9) illustrates these concepts:

9) $$\ldots,\ \boxed{\text{dass}}\ \underbrace{\underbrace{\text{wir den Dialog}}_{\text{RMF = 3}}\ \boxed{\text{fortführen}}\ \underbrace{\text{mit Russland}}_{\text{AF = 2}}}_{\text{TMF = 5}}$$

[...$\boxed{\text{that}}$ we the dialog $\boxed{\text{continue}}$ with Russia]

(EPICG_20080810_Formal sitting1_Ingrid Betancourt_I_de)

Here, the extraposed prepositional phrase in the after *field mit Russland* could have been placed in the middle field, therefore creating a theoretical middle field of 5 words.

All cases in which the number of words in the RMF is smaller than the number of words in the TMF are thus cases of extraposition, as in (10). If the RMF equals the TMF, there is no extraposition, as in (11), regardless of how many items the after field contains.

10) $$\ldots\text{dus ik hoop}\ \boxed{\text{dat}}\ \underbrace{\underbrace{\text{dit programma / dan / in werking}}_{\text{RMF = 5}}\ \boxed{\text{kan treden}}\ \underbrace{\text{in januari 2009}}_{\text{AF = 3}}}_{\text{TMF = RMF + AF = 8}}$$

[...so I hope $\boxed{\text{that}}$ this program then into force $\boxed{\text{can enter}}$ in January two thousand nine]

(EPICG_20082010_Erasmus Mundus_Marielle de Sarnez_I_nl)

11) $$\ldots\text{problemen}\ \boxed{\text{die}}\ \underbrace{\text{op dit moment geen oplossing}}_{\text{RMF = TMF = 5}}\ \boxed{\text{vinden}}$$

[...problems $\boxed{\text{that}}$ at this moment no solution $\boxed{\text{find}}$]

(EPICG_20082110_EU-Russia relations_Jean-Pierre Jouyet_I_nl)

2.2.2. Annotations

To gain deeper insights from the data, we annotated them with additional features. Data on speaker and interpreter sex were drawn from the EPICG metadata to check whether sex differences in memory skills had an effect on the length of the verbal brace in interpreting, as compared to non-mediated speech. In the corpus metadata, sex was determined solely on the basis of perceived voice characteristics. According to most research, starting with Lass, Hughes, *et al.* (1976), this method is reliable: listeners are able, on the basis of vowels pronounced in isolation, to identify speaker sex with an accuracy of over 95%.

Information on the structural properties of the source clauses was also included. Evidently, if the position of the verb in interpreted speech is determined by memory management, the position of the equivalent verb or clausal constituent in the source text is an important feature to take into account. As said before, French is an SVO language and all major constituents follow the verb. The cognitively least challenging solution for an interpreter is to produce speech that follows the source word order as closely as possible, within the limits of the target grammar. A French source is therefore a potential trigger for extraposition of clausal constituents in German and Dutch interpreted speech. We examined to what extent this was the case in our data by analyzing whether the structure of the target clause ran parallel to that of the source clause in each occurrence of extraposition. Example (12) illustrates a case in which source and target structures are parallel; while in (13) the extraposed constituent does not have an equivalent in the source.

12) ...nous n’avons pas atteint les performances que nous souhaitions
[...we have not achieved the performances that we wished]
German interpretation: ..., dass wir nicht die Leistungen erreicht haben die wir wünschten
[...that we not the performances have achieved that we wished]
(EPICG_2008.08.10_Arctic governance_Michel Rocard_I_nl)

13) ...que bien des fois vous avez senti peut-être la frustration
[...that quite often you have felt maybe frustration]
German interpretation: ...dass Sie häufig frustriert waren darüber
[...that you often frustrated were about that]
(EPICG_20080810_formal sitting1_Ingrid Betancourt_I_de)

Evidently, examples such as (12) are more informative for memory strategies than (13). If extrapositions are triggered by the structural properties of the source clause, it could mean that interpreters to a certain extent sacrifice target acceptability to memory management: they deliver the verb closer to the position it occupies in the source text, even though this is not the most typical position in the target language. However, causality is impossible to establish in these cases: on the basis of our data we cannot tell whether interpreters extrapose because they are triggered by the source or for other reasons.

3. Results

In the following sub-sections the results of the corpus analyses will be presented and discussed. We first compare non-mediated language and interpreting and then focus on interpreting.

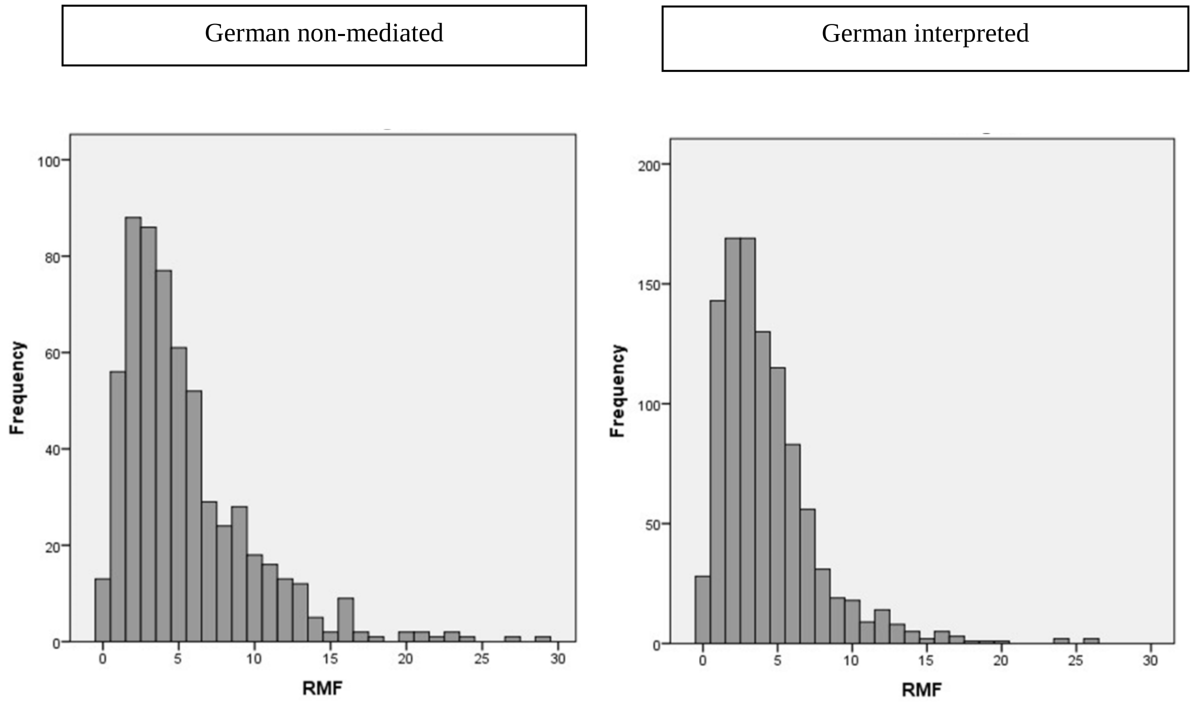
3.1. Length of the RMF

Table 2 shows the results for the length of the real middle field in German non-mediated speech and interpreted speech, i.e. the number of words between the subordinating conjunction and the verbal group. The results are presented as averages per speaker, as well as for all subordinates. It appears that interpreters into German overall have a shorter RMF, both in terms of mean number of words and in terms of the median. The interpreters also present less variation than original speakers as a group.

Table 2. Real middle field length: German

	N	Mean RMF	Median RMF	Standard Deviation	Minimum	Maximum
Per speaker						
German non-mediated	39	5.69	5.56	1.806	2.56	10.50
German interpreted	41	4.29	4	2.785	0.04	14.18
For all subordinates						
German non-mediated	602	5.41	4	4.291	0	29
German interpreted	1014	4.27	3	3.366	0	26

To check whether this difference is statistically significant, a visual inspection of the histograms in Figure 1 was carried outfor all subordinates to determine what kind of statistical test would be appropriate.



German non-mediated

German interpreted

Figure 1. RMF frequencies in German data

As the distribution for both groups is strongly skewed to the right, a non-parametrical Mann-Whitney U Test was performed on the data per speaker. The difference between original speakers and interpreters appears to be statistically highly significant (U=491.5; p=0.003). This suggests that interpreting may have an impact on the memory efforts the German interpreters are prepared to spend: on average the RMF of interpreters is 21% or one full word shorter than that of original German speakers, with a mean length of 4.9 words as opposed to 5.69 words.

For Dutch, the data are presented in Table 3. Again it appears that interpreters use shorter RMFs than original speakers, both in terms of mean length and in terms of the median. In the interpreted data, the mean RMF is 19% or almost a word shorter. As a group, interpreters also appear to act more homogeneously than original speakers.

Table 3. Real middle field length: Dutch

	N	Mean RMF	Median RMF	Standard Deviation	Minimum	Maximum
Per speaker						
Dutch non-mediated	24	4.48	4.25	0.904	3	6.67
Dutch interpreted	39	3.60	3.57	0.990	0.38	5.48
For all subordinates						

Dutch non-mediated	659	4.49	4.00	3.519	0	22
Dutch interpreted	1185	3.63	3.00	3.207	0	34

As was the case for German, the distribution for both groups for all subordinates is strongly skewed to the right, requiring non-parametrical testing to determine whether the differences are significant.

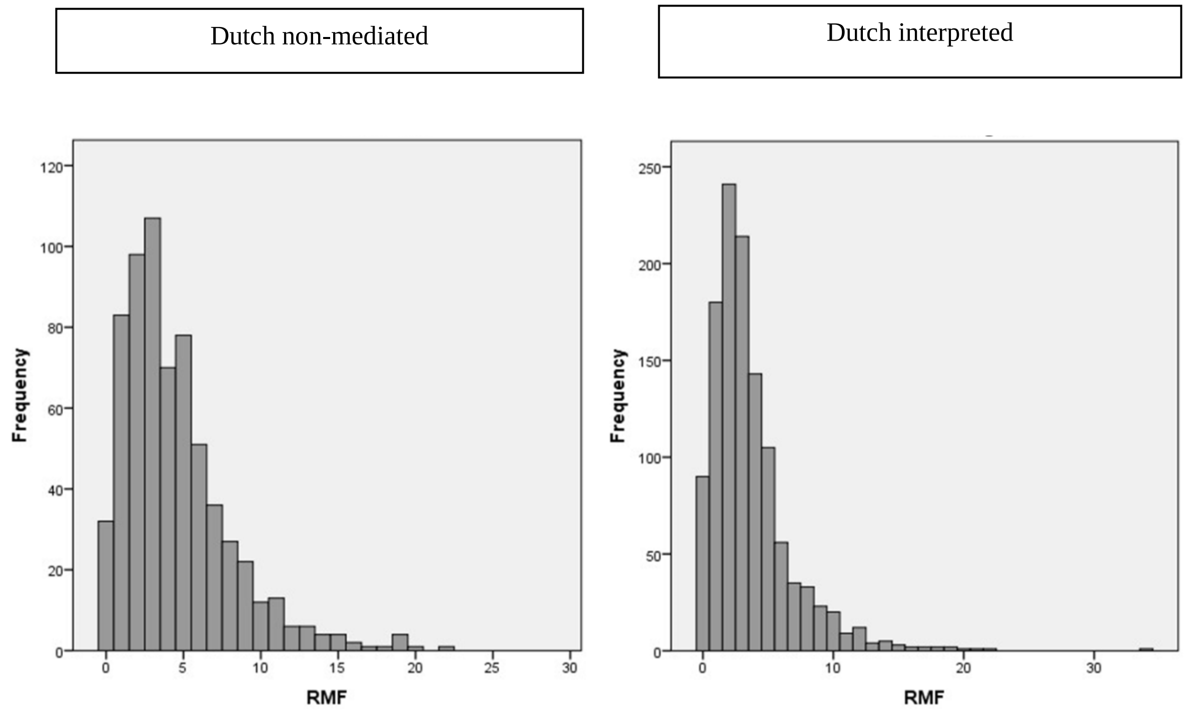


Figure 2. RMF frequencies in Dutch data

The Mann-Whitney U Test shows that the difference between Dutch interpreters and original speakers is statistically highly significant (U=235; p=0.001).

As the results point in the same direction for both languages and the differences are in both cases highly significant, the tendency in interpreters to have a shorter RMF may be said to be robust. Furthermore, in both languages interpreted RMFs are more homogeneous than non-mediated RMFs, even though there are more data for interpreters than for original speakers. This may indicate either that the shorter RMF is the result of a strategy interpreters commonly apply or that some sort of maximum memory capacity is reached, beyond which it becomes increasingly difficult to organize a verbal brace.

It also appears from Tables 2 and 3 and from Figure 3 (drawn from all subordinates) that in both interpreted and non-mediated speech, German RMFs are longer than Dutch RMFs, confirming a tendency documented in the literature.

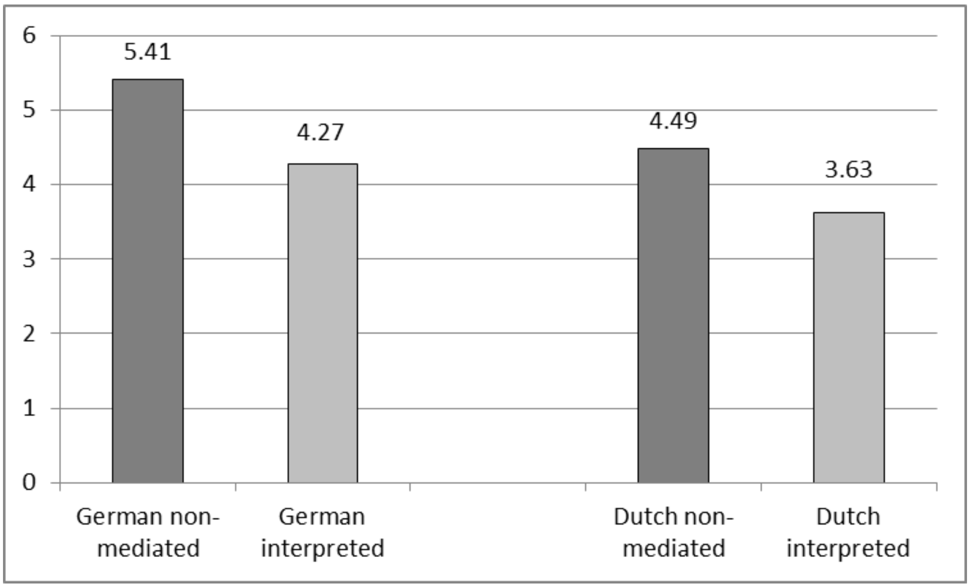


Figure 3. Mean RMF length in both varieties of German and Dutch

The Mann-Whitney U test was performed on the data per speaker in order to determine whether there is a difference between German and Dutch RMF. The difference is statistically highly significant for original speakers (U=261; p=0.003) but not for interpreters (U=633; p=0.262). It is noteworthy in this respect that the mean RMF in interpreted German is shorter than in Dutch non-mediated speech.

Sex was reported to be an important extra-linguistic parameter in the positioning of the verb: on average, women tend to use the middle field more than men for prepositional phrases, making the middle field longer. The data from our sub-corpora, presented in Tables 4 and 5, confirm this general trend for German original speakers, but not for Dutch original speakers and interpreters.

Table 4. Sex differences in RMF length in German

	N	Mean RMF	Median RMF	Standard Deviation	Minimum	Maximum
Female non-mediated	23	6.23	6.17	1.992	3.36	10.50
Male non-mediated	16	4.93	5.28	1.176	2.56	6.67
Female interpreted	33	4.05	4.00	2.980	0.04	14.18
Male interpreted	8	5.26	4.75	1.456	3.69	7.78

Table 5. Sex differences in RMF length in Dutch

	N	Mean RMF	Median RMF	Standard Deviation	Minimum	Maximum
Female non-mediated	7	4.61	4.66	1.060	3.27	6.44
Male non-mediated	17	4.43	4.21	0.862	3	6.67
Female interpreted	20	3.72	3.53	0.941	2.36	5.48
Male interpreted	19	3.47	3.57	1.049	0.38	5.15

In both languages, mean RMF is longer in female non-mediated speech than in male. For German, this difference is significant (U=110.5; p=0.035), but not for Dutch (U=55; p=0.804). In interpreted speech,the difference is not statistically significant in German (U=88; p=0.155), nor in Dutch (U=178; p=0.749).

There is no straightforward way to account for these observations. First, our findings for German non-mediated data are supported by trends observed in other types of non-mediated speech and can easily be related to observed differences in memory skills. On the other hand, in activities which draw heavily on memory, such as interpreting, the female advantage seems to disappear. The difference between non-mediated data in German and Dutch could be partly explained by the fact that German middle field is longer than Dutch middle field for these data. Indeed, with shorter middle fields, the female advantage in memory capacity might not be exploited.

3.2. Extraposition

It has become clear that interpreters tend to unload their memories more than original speakers when it comes to producing verbal braces. The question that we try to answer in this section is whether they do so by means of extraposition. In other words, are the shorter RMFs the result of extraposition as a strategy to unload memory, or are the structures that interpreters produce just shorter overall?

Table 6 below shows the frequencies of structures with extraposition and structures without extraposition per language and speech type.

Table 6. Frequencies of extraposition per language and speech type

	N	Structures with extraposition	Structures without extraposition
German non-mediated	39	2.15 (14%)	13.23 (86%)
German interpreted	41	5.49 (22%)	18.98 (78%)
Dutch non-mediated	24	9.08 (33%)	18.50 (67%)
Dutch interpreted	39	10.44 (35%)	19.79 (66%)

It appears that in all varieties, extraposition is the exception rather than the rule. Structures ending with the verb are at least twice as frequent as structures with extraposed constituents. As expected, extraposition is more popular in Dutch than in German. There is no significant difference between interpreters and original speakers in the use of extraposition in German (U = 718.5; p=0.425) nor in Dutch -(U = 427; p=0.565).

The shorter RMF found in interpreting both for Dutch and German therefore does not seem to be explained by a higher frequency of extraposition.

Finally, we checked whether extraposition is subject to sex differences. The results are shown in Tables 7 and 8.

Table . Sex differences in extraposition: German

	N	Cases of extraposition	Cases of no extraposition
Female non-mediated	23	1.43 (11%)	11.61 (89%)
Male non-mediated	16	3.19 (17%)	15.56 (83%)
Female interpreted	33	5.55 (23%)	18.97 (77%)
Male interpreted	8	5.25 (22%)	19 (78%)

Table 8. Sex differences in extraposition: Dutch

	N	Cases of extraposition	Cases of no extraposition
Female non-mediated	7	11.29 (33%)	22.43 (67%)
Male non-mediated	17	8.18 (33%)	16.88 (67%)

Female interpreted	20	12.25 (34%)	23.40 (66%)
Male interpreted	19	8.53 (35%)	16.00 (65%)

As we can see, the percentages are similar between men and women, for all groups, and the Mann-Whitney U tests show no statistically significant differences, for German non-mediated speech (U=123, p=0.084) and interpreted (U=80.5, p=0.091), nor for Dutch non-mediated speech (U=44, p=0.349) and interpreted (U=158.5, p=0.380). A possible explanation may lie in the intensive memory training interpreters are exposed to during their training and their daily practice, which is likely to close the gap between the sexes.

3.3. Influence of the source structure

In this last section, we check to what extent the target word order in interpreting, extraposition in particular, is triggered by the structural properties of the source clause. If so, this would be evidence that interpreters prioritize memory management over target acceptability. Since structures with extraposition in an SOV language are structurally more similar to an SVO word order than structures without extraposition, the cognitively least demanding way to interpret SVO into SOV is to maximize extraposition. However, extraposition is not the preferred option in the target language and therefore raises acceptability issues. Table 9 presents the frequencies of extraposition in clauses that are structurally similar and dissimilar to the corresponding source clauses.

Table 9. Structural similarity of target extraposition and source clause word order

	German interpreted	Dutch interpreted
Structurally similar to source	201 (89.3%)	369 (90.4%)
Structurally dissimilar to source	23 (10.2%)	34 (8.3%)
Undetermined	1 (0.4%)	5 (1.2%)
Total	225	408

In both languages, about 90% of extrapositions occur in contexts where they are potentially triggered by the word order in the source clause. This seems to suggest that interpreters do indeed use extraposition as a tool for memory management, following the source word order as closely as possible.

4. Conclusions

The aim of this study was to assess whether the increased cognitive effort of simultaneous interpreting had an influence on the position of the verb in SOV target languages. In an SOV language, the verb is placed at the end of the clause structure. In simultaneous interpreting from an SVO language into an SOV language, the verb therefore has to be stored in memory. The longer the middle field, the more effort is required to keep the verb stored. Given the additional cognitive load in interpreting, we hypothesized that interpreters would generally try to deliver the verb sooner than original speakers, keeping the middle field short and the memory effort as low as possible. Corpus data from EPICG and a collection of German transcriptions made specifically for this study supported this hypothesis for the two target languages involved (German and Dutch from French source speeches).

To find out whether the shorter middle field in interpreting is the result of a deliberate attempt to extrapose clausal constituents, we also checked whether extraposition is more frequent in interpreted than in non-mediated speech. No differences were found for both languages.. This means that the middle field in interpreting is not kept short by a more intensive use of extraposition, i.e. delaying constituents and advancing the delivery of the verb. However, other factors such as a general tendency towards simplification in spoken language could have an influence on the use of extraposition and more research is needed in order to enable a clearer understanding of these results, find other potential factors, as well as to confirm the reliability of the chosen parameters.

We also hypothesized that, given the well-documented female advantage in memory skills and their presumably resulting tendency to fill up the middle field, female interpreters and female original speakers would both maintain longer middle fields than males. Although our results for German non-mediated speech confirmed this general tendency, the data for Dutch and the German interpreted data showed no sex differences. The exception found for German non-mediated speeches might be linked to the longer middle fields found in German compared to Dutch. Indeed the female advantage might only be visible when the memory is at risk of being overloaded. When it comes to the use of extraposition, no differences were found.A possible explanation may lie in the intensive memory training interpreters are exposed to during their training and their daily practice.

Finally, it was also examined whether extraposition was favoured by similarities in clause word order to that of the source. This appeared the case in a vast majority of the occurrences, providing evidence that interpreters use extraposition as a tool in memory management insofar as it enables them to stay as close as possible to the source word order.

PART 3: CONCLUSION

Chapter 1. Overview of findings

This research was based on an assumption rarely explored in interpreting studies: compared to male interpreters, female interpreters need to dedicate less attentional resources to the concurrent cognitive processes involved simultaneous interpreting. This assumption might seem quite outdated given the general trend towards the abolishment of sex-based discrimination. Indeed, assumptions of a female advantage for the interpreting task were formulated thirty years ago (Gran & Fabbro 1988). However, despite advances in research into simultaneous interpreting, no conclusive findings have been reached on the potential role of sex in the cognitive processes of interpreting. This is mostly due to a clear lack of research including the interpreter's sex as a predictor.

Moreover, our assumption is not based on the same findings as Gran & Fabro's assumptions. Indeed, they claimed that women had a more symmetrical use of both hemispheres when dealing with language, which gave them an edge for tasks like simultaneous interpreting. However, this claim has mostly been refuted today. Our assumption has been formulated based on findings that show a female advantage for cognitive tasks linked to simultaneous interpreting: speech production and comprehension, as well as verbal memory. While these findings have not been refuted (as yet), they are not entirely conclusive as most effect sizes are small. This lack of conclusive findings on potential cognitive differences is even more reasons to explore the role of sex in one specific task, i.e. simultaneous interpreting, in order to increase data on the topic.

Accordingly, the cognitive processes in interpreting were operationalised into four different research topics with their own hypothesis. These topics are mostly linked to the way interpreters manage their memory capacity and respond to cognitive load. Most specifically, the management of memory capacity is closely linked to the length of Ear-Voice span and the use of extraposition. Indeed, the length of EVS is limited by interpreters' working memory capacity (Gile, 1999) and the longer the EVS, the more working memory capacity is required. Moreover, extraposition is considered a way of preventing the memory capacity from being overloaded. Interestingly, in both studies, sex has no influence, while it is believed that women perform better in verbal memory tasks. It was assumed that the length of the EVS would be shorter for women, because they need to dedicate fewer cognitive resources to the interpreting task, and the EVS length is positively influenced by the intensity of the cognitive processes. This hypothesis was refuted as no sex differences were found in the EVS. When it comes to the influence of predictors, it was discovered that the EVS increases when the number of uttered disfluencies (lengthenings, false starts and filled pauses) by the interpreter increases, which seems logical as disfluencies delay the output of the interpreter and therefore increase EVS. The EVS also increases when the original speaker's delivery rate increases, which is in line with previous findings (De Groot, 1997; Gerver, 1969; Lee, 2002), showing that when the rate of input the interpreter needs to process increases, the cognitive effort increases and so does the EVS. Contrary to Lee's findings (2002), the length of segments between pauses produced by the interpreter and the interpreter's delivery rate seem irrelevant. However, several predictors did not have a significant influence on the EVS. It appears clearly that silent pauses have no influence on the EVS, contrary to Lee's findings (2002). Moreover, unlike Barik's findings (1973), we did not find a significant influence of the type of delivery on the EVS. For the last corpus study, it had been discovered that on average, women tend to have a longer middle field than men. While our data confirm this general trend for German original speakers, they do not for interpreters and for Dutch.

The third research topic, the production of disfluencies, is closely linked to the management of cognitive load, as well as to speech production. Indeed, disfluencies are believed to be the consequence of cognitive load. Moreover, since women are believed to perform better at several tasks linked to speech production, more specifically verbal fluency, we had assumed that they would produce fewer disfluencies. This hypothesis was mostly confirmed by our data, since male interpreters were found to produce more disfluencies, especially false starts and lengthening. Interestingly, this topic was also the most represented in studies on sex differences, where similar findings had been yielded (Shriberg 1996; Bortfeld 2001). The EVS has also been found to increase the production of most disfluencies, as well as the source speaker's delivery rate and the interpreter's number of filled pauses. Other predictors are negatively associated with all types of disfluencies: interpreter's delivery rate and duration of silent pauses and the source speaker/interpreter speaking time ratio.

Finally, the fourth corpus study, the rendition of numbers, is actually linked to all cognitive processes involved in interpreting instead of a specific one. Indeed, given the difficulty that numbers represent for interpreters, they need to have sufficient resources available when dealing with them. Moreover, prevent interpreters from using common strategies (e.g. anticipation) which usually help preserve working memory capacity. Our hypothesis was that women have more cognitive resources available when dealing with the rendition of numbers and will therefore render more numbers than men. Moreover, given that men have fewer cognitive resources available when dealing with numbers, we hypothesize that they would resort to one of the strategies applied by professionals more often, i.e. present a shorter EVS. Both hypotheses were refuted by our data. Our findings confirm that the shorter the EVS, the more accurate the rendition of numbers , and that a high input rate increases the number of omissions. However, results find no significant influence of the nature and complexity of numbers and source and target languages on the rendition of numbers.

This research project has also allowed to analyse several variables in real-life settings, while they had mostly been studied in experimental designs. Our findings indicate that the median value for EVS in real-life settings is 3 seconds, which is in line with previous studies. EVS lengths range between -1.56 and 16.87 seconds and uttered anticipations account for about 0.1% of the data, which is much lower than some previous findings. However the values for the input rate are much higher than the values usually applied in experiments. Indeed the mean input rate is 156 words per minute, while the ideal rate is considered to be 120 words per minute. When it comes to the interpreter's rate, the mean value is 144 words per minute. While Gerver (1969) has observed a maximum rate of 110 words/min for interpreters. These high values are most probably linked to the specific settings of the plenary sessions of the European Parliament were speakers often have short allocated speaking time. However, anecdotal evidence tends to show that a fast-speaking rate is becoming the norm. Our data indicate that the EVS increases when the input rate increases. Moreover, the frequency of disfluencies produced by the interpreter appears to increase when the input rate increases. As expected the frequency of errors and omissions in the rendition of numbers was much lower in our data compared to experimental data (a total of 21% in our data compared to about 40% for experiments), given that interpreters often receive help from their colleagues or the written speech. Our findings also confirm that omission is the most frequent category after complete rendition. The most frequent type of disfluencies for the interpreter is filled pauses. Our data show that the most frequent type of disfluency in both interpreting and spontaneous speech is the silent pause followed by the filled pause. Counting only these types of uttered disfluencies under study, about 7.8% of uttered tokens are disfluent in interpreting, compared to about 2.6% for original speakers. The frequency of disfluencies in our corpus therefore appears to be slightly lower than that recorded in the experimental study by Christodoulides & Lenglet (2014). The average total duration of silent pauses is very similar between original speakers and interpreters, but the number of pauses is higher in spontaneous speech, which means that interpreters produce fewer but longer pauses than original speakers.

When it comes to the influence of languages, our findings show that Dutch triggers the longest EVS, which is in line with findings indicating that SOV source languages require a longer EVS. Moreover, the source language French seems to trigger fewer disfluencies than English, which on the contrary goes against the literature according to which language pairs with a SOV constituent order in the source language increase cognitive load. Moreover, Dutch-speaking source speakers reduce their middle field, compared to their German-speaking counterparts, confirming a tendency documented in the literature.

Given that our data are based on a parallel corpus, several conclusions can also be reached on the differences and similarities between source speakers and interpreters. Our results show that interpreters produce more disfluencies than source speakers.While this tends to confirm that interpreting is cognitively more demanding than spontaneous speech, it also contradicts the idea that interpreters omit useless elements in the source speech. Interestingly, the interpreters have a higher speech/pause ratio, with a mean value of 7.14 than source speakers, with 5.82. The last corpus study shows that interpreters into Dutch and German tend to shorten their middle field more than original speakers and use extraposition partly because they imitate the clause word order of the source speech (in this case French).

Chapter 2. Next steps

While this dissertation has allowed for interesting findings, we think that this is only a beginning. Thanks to the methodologies applied in order to transcribe, time tag and annotate the present corpus, as well as to analyse and quantify the results (see Appendixes 1 and 2), we hope to inspire other researchers to use corpus data to observe cognitive processes in interpreting. More specifically we encourage researchers to include sex as a predictor variable in their experimental designs in order to confirm or refute our findings. We would also wholeheartedly welcome complementary experimental research on these topics to confirm or contradict our corpus-based findings in highly controlled circumstances.

Several potentially relevant predictors have not been included in our analysis and might yield interesting results. These variables have not been included mostly because they were difficult to operationalize and require a depth of analysis that could not be attained in the framework of this research. Among these variables, we think of: the length and complexity of sentences, the input and output lexical density and syntactic complexity. The topic or domain of speeches might also be taken into account. More specifically, the position of disfluencies in sentences could be significant. Moreover, some types of disfluencies were not taken into account, such as for instance repetitions, grammatical errors, articulation rate and alterations in voice quality. Moreover, one research topic could not be included to this dissertation due to time constraints: the rendition of lists of items by the interpreter. This topic would have complemented results on the rendition of numbers, as well as providing insights on omissions.

Moreover, the inclusion of other languages to the corpus would allow for more representative conclusions to be drawn. As an interpreter working from German, I sometimes have the feeling that the interpreting technique and strategies applied differ highly from other passive languages, like English for example. I therefore think that adding German to the whole corpus will greatly improve the quality of studies. .

We are particularly keen on sharing our corpus online, as well as the numerous tutorials and scripts developed for its creation and analysis. While this requires some technical support, it would greatly contribute to corpus-based research in interpreting by guaranteeing the replicability of findings. Finally, having access to the identities, or at least to some demographic information about the interpreters working at the European Parliament would allow for findings to be more rigorous. Moreover, including the interpreters in the research effort would also allow for the interpreter's personal preferences and strategies, level of preparation and experience to be taken into account.

Bibliography

Acheson, D. & MacDonald, M. (2009). Twisting tongues and memories: explorations of the relationship between language production and verbal working memory. *Journal of Memory and Language*, *60*, 329-350.

Aerts, A., van Mierlo, P., Hartsuiker, R., Hallez, H., Santens, P. & De Letter, M. (2013). Neurophysiological Investigation of Phonological Input: Aging Effects and Development of Normative Data. *Brain and Language, 125*(3), 253–263.

Ahrens, B. (2005). Prosodic phenomena in simultaneous interpreting: A corpus-based analysis. *Interpreting, 7*(1), 51–76.

Ahrens, B., Kalderon, E., Krick C., & Reith W. (2010). fMRI for exploring simultaneous interpretation. In D. Gile, G. Hansen & N. Pokorn (Eds.), *Why Translation Studies Matter.*Amsterdam: Benjamins, 237-250.

Albert, M. & Obler, L. (1978) *The bilingual brain: Neuropsychological and neurolinguistic aspects of bilingualism*. New York: Academic Press.

Aleman, A., Kahn, R.S. & Selten, J. P. (2003). Sex differences in the risk of schizophrenia: evidence from meta-analysis. *Archives of General Psychiatry, 60*, 565–571.

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

Angelili, C. (2004). Revisiting the interpreter's role: a study of conference, court, and medical interpreters in Canada, Mexico, and the United States. Amsterdam: John Benjamin.

Archer, J. (2004). Sex differences in aggression in real-world settings: a meta-analytic review. *Review of General Psychology, 8*(4), 291–322.

Arnold, J., Fagnano, M. & Tanenhaus, M. (2003). Disfluencies signal theee, um, new information. *Journal of Psycholinguistic Research, 3*, 25–36.

Arnold, J., Tanenhaus, M., Altmann, R. & Fagnano, M. (2004). The old and thee, uh, new: Disfluency and reference resolution. *Psychological Science, 15*, 578–582.

Arnold, J., Wasow, T., Losongco, A. & Ginstrom, R. (2000). Heaviness vs. newness: The effects of structural complexity and discourse status on constituent ordering. *Language, 76*(1), 28–55.

Atkinson, R. C. & Shiffrin, R.M. (1971, April 19). The control processes of short-term memory. *Scientific American, 225(2)*, 82-90.

Bachy, S., Dister, A., Francard, M., Geron, G., Giroul, V., Hambye P., Simon, A. C. & Wilmet, R. (2007, April 18). Conventions de transcription régissant les corpus de la banque de données VALIBEL. Retrieved from https://www.uclouvain.be/cps/ucl/doc/valibel/documents/conventions_valibel_2004.pdf.

Baddeley, A. (1996). Exploring the Central Executive. The Quarterly Journal of Experimental Psychology Section A, 49(1), 5–28.

Baddeley A.D. & Hitch G. (1974). Working Memory. *Psychology of Learning and Motivation, 8*, 47-89.

Baer, B. J. & Massardier-Kenney, F. (2016). Gender and sexuality. In C. V. Angelelli & B. J. Baer (Eds.), *Researching Translation and Interpreting*. London: Routledge, 83-96.

Baes, E. (2012). *Analyse des différences entre interprètes masculins et féminins au Parlement européen en termes de décalage*. [Analysis of the differences in décalage between male and female interpreters at the European Parliament.] (Unpublished Master's thesis). Hogeschool Gent, Ghent, Belgium.

Díaz-Galaz, S., Padilla, P., & Bajo, M. T. (2015). The role of advance preparation in simultaneous interpreting: A comparison of professional interpreters and interpreting students. *Interpreting, 17(1),* 1-25.

Bajo, V. M., Nodal, F.R., Moore, D.R., King, A.J (2010). The descending corticocollicular pathway mediates learning-induced auditory plasticity. *Nat Neurosci 13*,253–260.

Baker, C. (1993). Foundations of bilingual education and bilingualism. *Multilingualism Matters*, xvi + 319 pp.

Bakti, M. (2009). Speech disfluencies in simultaneous interpreting. In D. De Crom (Ed.), *(Trans)formation of identities: Selected papers of the CETRA research seminar in translation studies 2008*. Leuven: CETRA, 1-18.

Barbizet, J. (1968).«Les bases neuro-anatomiques de la genèse de la significationdans le langage oral» in Robert Husson et al. (éds), Mécanismes cérébraux du langage oral et structures des langues, Paris: Masson & Cie

Barik, H. C. (1971). A Description of Various Types of Omissions, Additions and Errors of Translation Encountered in Simultaneous Interpretation. *Meta, 16(4)*, 199–210.

Barik, H. C. (1973). Simultaneous interpretation: temporal and quantitative data. *Language and Speech, 16*(3), 237-270.

Barik, H. C. (1975). Simultaneous interpretation: qualitative and linguistic data. *Language and Speech, 18*(3), 272–297.

Bernardini, S., Ferraresi, A., Russo, M., Collard, C. & Defrancq B. (2018). Building Interpreting and Intermodal Corpora: A How to for a Formidable Task. In M. Russo, C. Bendazzoli & B. Defrancq (Eds.), *Making Way in Corpus-based Interpreting Studies*. Singapore: Springer, 21-42.

Bendazzoli, C. & Sandrelli, A. (2005, May). An Approach to Corpus-Based Interpreting Studies: Developing EPIC (European Parliament Interpreting Corpus). In H. Gerzymisch-Arbogast & S. Nauert (Eds.), *Proceedings of the Marie Curie Euroconferences MuTra: Challenges of Multidimensional Translation*. Saarbrücken.

Bermingham, D., Hill, R. H., Woltz, D. & Gardner M. K. (2013). Cognitive strategy use and measured numeric ability in immediate- and long-term recall of everyday numeric information. *PLoS One 8*(3), e57999.

Bernardini, S., Ferraresi, A., Russo, M., Collard, C. & Defrancq B. (2018). Building interpreting and intermodal corpora: A how to for a formidable task. In C. Bendazzoli, B. Defrancq & M. Russo (Eds.), *Making Way in Corpus-based Interpreting Studies*. Singapore: Springer, 21-42.

Betz, S. & Wagner P. (2016). Disfluent Lengthening in Spontaneous Speech. In O. Jokisch (Ed.), *Elektronische Sprachsignalverarbeitung (ESSV)*, *Studientexte zur Sprachkommunikation*. Dresden: TUD Press, 135–144.

Betz, S., Wagner, P. & Schlangen, D. (2015). Micro-Structure of Disfluencies: Basics for Conversational Speech Synthesis. In *Proceedings of Interspeech* (pp. 2222-2226). Dresden, Germany.

Bevilacqua, L. (2009). The position of the verb in germanic languages and simultaneous interpretation. *The Interpreters' Newsletter, 14*, 1-31.

Bleecker, M. L., Bolla-Wilson, K. , Agnew, J. & Meyers, D. A. (1988), Age-related sex differences in verbal memory. *J. Clin. Psychol., 44*, 403-411.

Bock, K. & Levelt, W. (1994). Language production. Grammatical encoding. In M.A. Gernsbacher (Ed.), *Handbook of psycholinguistics*. San Diego: Academic Press, 945-84.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International 5*(9), 341-345.

Bolla-Wilson, K. & Bleecker, M. (1986). Influence of verbal intelligence, sex, age, and education on the Rey Auditory Verbal Learning Test. *Developmental Neuropsychology, 2*(3), 203–211.

Boomer, D. (1965). Hesitation and grammatical encoding. *Language and Speech, 8*(3), 148-58.

Born, M. P., Bleichrodt, N. & van der Flier, H. (1987). Cross-cultural comparison of sex related differences on intelligence tests. *Journal of Cross-Cultural Psychology, 18*, 283-314.

Bortfeld, H., Leon, S., Bloom, J., Schober, M. & Brennan, S. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech*, *44*(2), 123–147.

Braecke, C. (1990). Uit de tang' of [+ or -] prominent? *Taal en Tongval, 3*, 125-134.

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

Brennan, S. & Schober, M. (2001). How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Language, 44*, 274–296.

Broadbent, D. (1958). *Perception and Communication*. London: Pergamon Press.

Bührig, K., Kliche, O., Pawlack, B. & Meyer, B. (2012). The corpus "Interpreting in Hospitals": Possible applications for research and communication training. In T. Schmidt & K. Wörner (Eds.), *Multilingual Corpora and Multilingual Corpus Analysis. Hamburg Studies in Multilingualism*. Amsterdam: John Benjamins, 305-315.

Burnard, L. (2002). A retrospective look at the British National Corpus. In B. Kettemann & G. Marko (Eds.), *Language and Computers: Studies in Practical Linguistics*. Amsterdam/New York: Rodopi, 51-70.

Cahill, L. (2017). An issue whose time has come. *Journal of Neuroscience Research, 95*, 12–13.

Camarata, S., & Woodcock, R. (2006). Sex differences in processing speed: Developmental effects in males and females. *Intelligence*, *34*, 231–252.

Cecot, M. (2001). Pauses in simultaneous interpretation: a contrastive analysis of professional interpreters' performances. *The Interpreters' Newsletter, 11*, 63-85.

Chambers, J. K. & Trudgill, P. (1998). *Dialectology* (2nd ed.). Cambridge, UK: Cambridge University Press.

Chesterman A. (1997). *Memes of Translation: The Spread of Ideas in Translation Theory*. Amsterdam/Philadelphia: John Benjamins.

Chernov, G. (1978). *Theory and Practice of Simultaneous Interpreting*. Moscow: Mezhdunarodnye otnosheniya.

Chiarello, C., Welcome, S.E., Halderman, L.K., Towler, S., Julagay, J., Otto, R.& Leonard, C.M. (2009).A large-scale investigation of lateralization in cortical anatomy and word reading: are there sex differences? *Neuropsychology, 23*, 210–222.

Chmiel, A. (2015). Przetwarzanie w tłumaczeniu symultanicznym. [Simultaneous interpretation processing.] In A. Chmiel & P. Janikowski (Eds.), *Dydaktyka tłumaczenia ustnego* [*Didactics of interpreting*]. Katowice, Poland: Stowarzyszenie Inicjatyw Wydawniczych, 227–247.

Chmiel, A. & Mazur, I. (2013). Eye tracking sight translation performed by trainee interpreters. In C. Way, S. Vandepitte, R. Meylaerts & M. Bartlomiejczyk (Eds.), *Tracks and Treks in Translation Studies: Selected papers from the EST Congress* (pp. 189-205). Leuven, Belgium

Christodoulides, G. (2013). Prosodic features of simultaneous interpreting. In P. Mertens & A. Simon (Eds.), *Proceedings of the Prosody – Discourse Interface Conference* (pp. 33–37). Leuven, Belgium.

Christodoulides, G. & Lenglet, C. (2014). Prosodic correlates of perceived quality and fluency in simultaneous interpreting. In *Proceedings of the 7th Speech Prosody Conference* (pp. 1002-1006). Trinity College, Dublin, Ireland.

Christoffels I. K., de Groot A. M. B., Waldorp L. J. (2003). Basic skills in a complex task: a graphical model relating memory and lexical retrieval to simultaneous interpreting. Biling. (Camb. Engl.) 6, 201–211

Christoffels et al ; 2006 Memory and language skills in simultaneous interpreters: The role of expertise and language proficiency Journal of Memory and Language 54(3):324-345 ·

Christoffels, I. K., & De Groot, A. M. (2004). Components of simultaneous interpreting: Comparing interpreting with shadowing and paraphrasing. *Bilingualism: Language and Cognition, 7*(3), 227–240.

Clark, H. & Fox Tree, J. (2002). Using uh and um in spontaneous speaking. *Cognition, 84*, 73–111.

Coates, J. (1993). *Women, men and language* (2nd ed.). London: Longman.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Collard, C. & Defrancq, B. (2017). Sex Differences in Simultaneous Interpreting: A Corpus-Based Study. Poster presented at the CIUTI Forum 2017.

Collard, P., Corley M., MacGregor L. & Donaldson D. (2008). Attention orienting effects of hesitations in speech: Evidence from ERPs. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 696–702.

Corley, M. & Hartsuiker, R. (2011). Why um helps auditory word recognition: the temporal delay hypothesis. *PLoS One, 6*(5), 197-192.

Corley M., MacGregor L. & Donaldson D. (2007). It's the way that you, er, say it: Hesitations in speech affect language comprehension. *Cognition, 105*, 658– 668.

Costa, P.T., Terracciano, A. & McCrae, R.R. (2001). Gender differences in personality traits across cultures: robust and surprising findings. *Journal of Personality and Social Psychology, 81*, 322–331.

Coussé, Evie (2009, May). *Changing word order in Dutch.* Paper presented at the 5th International Conference on Language Variation in Europe. København, Denmark.

Daro, V. & Fabbro, F. (1994). Verbal memory during simultaneous interpretation: effects of phonological interference. *Applied Linguistics, 15*, 365-381.

Damasio, A. R. (1992). The selfless consciousness. *Behavioral and Brain Sciences 15 (2)*,208-209.

Daneman, M. (1991). Working memory as a predictor of verbal fluency. Journal of *Psycholinguistic Research,20*, 445–464.

Daneman, M. & Carpenter, P. A. (1983). Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology: Learning. Memory. and Cognition, 9*, 561-583.

Deary, I. J., Irwing, P., Der, G. & Bates, T. C. (2007). Brother–sister differences in the g factor in intelligence: analysis of full, opposite-sex siblings from the NLSY1979. *Intelligence, 35*, 451–456.

Deary, I.J., Strand, S., Smith, P. & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence, 35*, 13–21.

Defrancq, B. (2013, September). *Women and men interpreting*. Paper presented at the first Talking to the World Conference, Newcastle, United Kingdom.

De Groot, A. M. (1997). The cognitive study of translation and interpretation. In J. H. Danks, G. M. Shreve, G.M. Fountain, & M. K. McBeath (Eds.), *Cognitive Processes in Translation and Interpreting.* London: Sage Publication, 25-56.

De Groot, A.M.B.& Christoffels, I.K.(2005). *Handbook of bilingualism: psycholinguistic approaches.* New York, United States : Oxford University Press, pp 327-330.

Desmet B., Defrancq B. & Vandierendonck, M. (forthcoming). Simultaneous interpretation of numbers and the impact of technological support. In C. Fantinuoli (Ed.), *Interpreting and Technology.* Berlin: Language science press.

De Schutter, G. (1994). Dutch. In E. König & J. van der Auwera (Eds.), *The Germanic languages*. London: Routledge, 439-477.

De Schutter, G. (2003). De zinsbouw in het vroege Middelnederlands. Een case study. *Taal en Tongval, 55*, 41-67.

De Sutter, G. & Van de Velde, M. (2008). Do the mechanisms that govern syntactic choices differ between original and translated language? A corpus-based translation study of PP extraposition in Dutch and German. In R. Xiao (Ed.), *Using corpora in contrastive and translation studies*. Newcastle: Cambridge Scholars Publishing, 144-163.

DeVries, G.J. (2004).Sex Differences in Adult and Developing Brains: Compensation, Compensation, Compensation. *Endocrinology, 145*,1063-1068.

Díaz-Galaz, S. (2011). The effect of previous preparation in simultaneous interpreting: Preliminary results. *Across Languages and Cultures, 12*(2), 173-191.

Díaz-Galaz, S., Padilla, P. & Bajo, M. T. (2015). The role of advanced preparation in simultaneous interpreting. A comparison of professional interpreters and interpreting students. *Interpreting, 17*(1), 1–25.

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

Donato, V. (2003). Strategies adopted by student interpreters in SI: A comparison between the English–Italian and the German–Italian language-pairs. *The Interpreters' Newsletter, 12*, 101–134.

Drach, E. (1937/1963). *Grundgedanken der deutschen Satzlehre*. Darmstadt: Wissenschaftliche Buchgesellschaft.

Duflou, V. (2016). Be(com)ing a conference interpreter: an ethnography of EU interpreters as a professional community. Amsterdam/Philadelphia, PA : John Benjamins.

Eisenberg, P. (1994). German. In E. König & J. van der Auwera (Eds.), *The Germanic Languages*. London: Routledge, 349-387.

Engelhardt, P., Corley, M., Nigg, J. & Ferreira, F. (2010). The role of inhibition in the production of disfluencies. *Memory and Cognition, 38*(5), 617–628.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102*, 211-245

Fantinuoli, C. (2018). The use of comparable corpora in interpreting practice and teaching. *The Interpreters' Newsletter ,23.*

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice, 40*(5), 532-538.

Ferreira, F., Lau, E. & Bailey, K. (2004). Disfluencies, language comprehension, and tree adjoining grammars. *Cognitive Science, 28*, 721–749.

Fox Tree, J. (1995). The effects of false starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory and Language, 34*, 709–738.

Fox Tree, J. (2001). Listeners' uses of um and uh in speech comprehension. *Memory and Cognition, 29*, 320–326.

Fox Tree, J. & Schrock J. (1999). Discourse markers in spontaneous speech: Oh what a difference an oh makes. *Journal of Memory and Language, 40*, 280–295.

Gale, S. D., Baxter, L., Connor, D. J., Herring, A. & Comer, J. (2007). Sex differences on the Rey Auditory Verbal Learning Test and the Brief Visuospatial Memory Test-Revised in the elderly: normative data in 172 participants. *Journal of Clinical and Experimental Neuropsychology, 29*(5), 561–567.

Garrett, M. (1975). Syntactic process in sentence production. In G. Bower (Ed.), *Psychology of learning and motivation: Advances in research and theory* (Vol.9). New York: Academic Press, 133-177.

Garcia-Landa, M.(1981). « La théorie du sens, théorie de la traduction et base de son enseignement ».In *l'enseignement de l'interprétation et de la traduction : de la théorie à la pédagogie*, Ottawa :University of Ottawa Press113-132

Gennatas, E. D., Avants, B. B., Wolf, D. H., Satterthwaite, T. D., Ruparel, K., Ciric, R., Hakonarson, H., Gur, R. E. & Gur, R. C. (2017). Age-related effects and sex differences in gray matter density, volume, mass, and cortical thickness from childhood to young adulthood. *Journal of Neuroscience, 37*, 5065–5073.

Gerritsen, M. & Stein, D. (1992). *Internal and external factors in syntactic change*. Berlin: Mouton De Gruyter.

Geffen, Gina, Moar, K. J., O'Hanlon, A. P., Clark, C. R., & Geffen, L. B.(1990). Performance measures of 16- to 86- year-old males and females on the auditory verbal learning test. *The Clinical Neuropsychologist, 4*, 45–63.

Gerver, D. (1969). The Effect of Source Language Presentation Rate on the Performance of Simultaneous Conference Interpreters. In F. Pöchhacker & M. Shlesinger (Eds.), *The Interpreting Studies Readers*. London/New York: Routledge, 52-66.

Gerver,D. (1975). A psychological approach to simultaneous *interpreting*. *Meta , 20*(2), 119-128.

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

Gile, D. (1997). Conference interpreting as a cognitive management problem. In J. Danks, G. Shreve, S. Fountain & M. McBeath (Eds.), *Cognitive Processes in Translation and Interpreting*. Thousand Oaks, CA: SAGE Publications, 196–214.

Gile, D. (1999). Testing the effort models' tightrope hypothesis in simultaneous interpreting – a contribution. *Hermes, 23*, 153-172.

Gile, D. (2008). Local cognitive load in simultaneous interpreting and its implications for empirical research. *Forum, 6*(2), 59–77.

Gile, D. (2009). Basic Concepts and Models in Translator and Interpreter Training. Amsterdam: Benjamins.

Gile, D. (2018, January 6) The effort models and gravitational model: Clarifications and update. Retrieved from http://www.cirinandgile.com/powerpoint/The-Effort-Models-and-Gravitational-Model-Clarifications-and-update.pdf.

Gillies, A. (2013). Conference Interpreting. A Student's Practice Book. Manchester: St. Jerome.

Goldman-Eisler, F. (1958). Speech production and the predictability of words in context. *Quarterly Journal of Experimental Psychology, 10*, 96–106.

Goldman-Eisler, F. (1967). Sequential Temporal Patterns and Cognitive Processes in Speech, *Language and Speech, 10*, 122-132.

Goldman-Eisler, F. (1972). Segmentation of input in simultaneous interpretation. *Journal of Psycholinguistic Research, 1*(2), 127-140.

Gósy, M. (2007). Disfluencies and Self-monitoring. *Govor, 26,* 91-110.

Gran, L. & Fabbro, F. (1988). The role of neuroscience in the teaching of interpretation. *The*

Interpreter's Newsletter, 1, 23 – 41.

Gur, R. E. & Gur, R.C. (2016). Sex differences in brain and behavior in adolescence: findings from the Philadelphia Neurodevelopmental Cohort. *Neuroscience & Biobehavioral Reviews*, 70, 159–170.

Guyette, T. & Baumgartner, J. (1988). Stuttering in the adult. In N. Lass, L. McReynolds, J. Northern & D. Yoder (Eds.), *Handbook of speech-language pathology and audiology*. Philadelphia, PA: B. C. Decker Inc, 640- 654.

Hodáková, S. (2009). Pamäť v simultánnom a konzekutívnom tlmočení. [Memory in simultaneous and consecutive interpreting.] Unpublished PhD dissertation. Konštantín Filozof University, Nitra.

Haeseryn, W. (1990). *Syntactische normen in het Nederlands. Een empirisch onderzoek naar volgordevariatie in de werkwoordelijke eindgroep* [Syntactic norms in Dutch. An empirical study on order variations in the verbal end group]. Nijmegen: KU Nijmegen.

Haeseryn, W. (1997). Achteropplaatsing van elementen in de zin. *Colloquium Neerlandicum, 13*, 303-326.

Haeseryn, W., Romijn, K., Geerts, G., de Rooi, J. & van den Toorn, M.C. (1997). *Algemene Nederlandse Spraakkunst*. Groningen/Deurne: Martinus Nijhoff – Wolters Plantyn.

Halpern, D. F. (2012). *Sex Differences in Cognitive Abilities* (4th ed.). New York, NY: Psychology Press.

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S. & Gernsbacher, M. A. (2007). The Science of Sex Differences in Science and Mathematics. *Psychological Science in the Public Interest, 8*(1), 1–51.

Harness, A., Jacot, L., Scherf, S., White, A. & Warnick, J. E. (2008). Sex differences in working memory. *Psychological Reports, 103*(1), 214-218.

Harrison, T. L., Shipstead, Z., Hicks, K. L., Hambrick, D. Z., Redick, T. S., & Engle, R. W. (2013). Working Memory Training May Increase Working Memory Capacity but Not Fluid Intelligence. *Psychological Science, 24(12)*, 2409–2419

Hartsuiker, R. & Kolk H. (2001). Error Monitoring in Speech Production: A Computational Test of the Perceptual Loop Theory (Statistical Data Included). *Cognitive Psychology, 42*(2), *113*(45).

Hawkins, J. (1994). *A performance theory of order and constituency.* Cambridge: Cambridge University Press.

Hedges, L. V. & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science, 269*(5220), 41-45.

Herlitz, A., Airaksinen, E. & Nordström, E. (1999). Sex differences in episodic memory: The impact of verbal and visuospatial ability. *Neuropsychology, 13*(4), 590–597.

Herlitz, A., Nilsson, L.-G. & Bäckman, L. (1997). Gender differences in episodic memory. *Memory and Cognition, 25*(6), 801–811.

Hervais-Adelman, A., Moser-Mercer, B ., & Golestani, N. (2011). Executive control of language in the bilingual brain: integrating the evidence from neuroimaging to neuropsychology. *Frontiers in Psychology, 2*, 234

Hervais-Adelman, A., Moser-Mercer, B., Michel, C. & Golestani, N. (2014). fMRI of SimultaneousInterpretation Reveals the Neural Basis of Extreme Language Control. *Cerebral Cortex, 25(12)*, 4727-4739.

Hervais-Adelman, A., Moser-Mercer, B. & Golestani, N. (2015). Brain functional plasticity associated with the emergence of expertise in extreme language control. *NeuroImage, 114*, 264-274.

Hickey, S. (2018). *The Female Perspective on the Feminisation of Conference Interpreting* (Unpublished Master's thesis). NUI Galway, Galway, Ireland.

Hines, M. (1990). Gonadal hormones and human cognitive development. In J. Balthazart (Ed.), *Brain and behaviour in vertebrates 1: Sexual differentiation, neuroanatomical aspects, neurotransmitters, and neuropeptides*. Basel, Switzerland: Karger, 51-63.

Hirnstein, M., Andrews L. & Hausmann M. (2014). Gender-stereotyping and cognitive sex differences in mixed- and same-sex groups. *Archives of Sexual Behavior, 43*, 1663–1673.

House, J., Meyer, B. & Schmidt, T. (2012). CoSi: A Corpus of Consecutive and Simultaneous Interpreting. In T. Schmidt & K. Wörner (Eds.), *Multilingual Corpora and Multilingual Corpus Analysis. Hamburg Studies in Multilingualism*. Amsterdam: John Benjamins, 295-304.

Howell P. & Young K. (1991). The use of prosody in highlighting alteration in repairs from unrestricted speech. *Quarterly Journal of Experimental Psychology, 43*, 733–758.

Hu, Yi & Anders Ericsson, K. (2012). Memorization and recall of very long lists accounted for within the long-term working memory framework. *Cognitive Psychology 64,* 235–66.

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist, 60*(6), 581-592.

Hyde, J. S. & Linn, M. C. (1988). Gender differences in verbal-ability - a meta-analysis. *Psychological Bulletin, 104*(1), 53–69.

Iliescu, D., Ilie, A., Ispas, D., Dobrean, A. & Clinciu, A.I. (2016). Sex differences in intelligence: a multi-measure approach using nationally representative samples from Romania. *Intelligence, 58*, 54–61.

Jäncke, L., Mérillat, S., Liem, F. & Hänggi, J. (2015). Brain size, sex, and the aging brain. *Human Brain Mapping, 36*, 150–169.

Jameson, A., Kiefer, J., Müller, C., Grossmann-Hutter, B., Wittig, F. & Rummer, R. (2009). Assessment of a user's time pressure and cognitive load on the basis of features of speech. In M. Crocker & J. Siekmann (Eds.), *Resource-adaptive cognitive processes.* Berlin: Springer, 171–204.

Jansen, F. (1978). Hoe krijgt een spreker zijn woorden op een rijtje? Taalgebruiksaspekten van de ‘PP over V’ konstruktie. In J. Kooij (Ed.), *Aspekten van de woordvolgorde in het Nederlands.* Leiden: Vakg, 70-104.

Jensen, A. R. & Reynolds, C. R. (1983). Sex differences on the WISC-R. *Personality and Individual Differences, 4*, 223–226.

Johnson, W., Carothers, A. & Deary, I. J. (2008). Sex differences in variability in general intelligence: a new look at the old question. *Perspectives on Psychological Science, 3*, 518–531.

Jones, R. (2002). Conference Interpreting Explained. Abingdon/New York: Routledge.

Kade, O., & Cartellieri, C. (1971). Some Methodological Aspects of Simultaneous Interpreting. *Babel, 17*(2), 12-16.

Kader, S., &. Seubert, S. (2014). Anticipation, Segmentation ... Stalling? How to Teach Interpreting Strategies. In D. Andres, & M. Behr (Eds.), *To Know How to Suggest... Approaches to Teaching Conference Interpreting* .Berlin: Frank & Timme, 125-144.

Kail, R. V. & Siegel, A. W. (1978). Sex and hemispheric differences in the recall of verbal and spatial information. *Cortex, 14*(4), 557-63.

Keith, T. Z., Reynolds, M. R., Patel, P. G. & Ridley, K. P. (2008). Sex differences in latent cognitive abilities ages 6 to 59: Evidence from the Woodcock-Johnson III tests of cognitive abilities. *Intelligence, 36*(6), 502-525.

Kimura, D. & Seal, B. (2003). Sex differences in recall of real or nonsense words. *Psychological Reports, 93*(1), 263-264.

Kinsbourne, M. & Cook, J. (1971). Generalized and lateralized effects of concurrent Verbalization on a unimanual skill. *Quarterly Journal of Experimental. 23*, 341-345.

Klapp, S., Greim, D., et al. (1981). Buffer storage of programmed articulation and articulatory loop: Two names for the same mechanism or two distinct components of short-term memory? In J. Long & A. Baddeley (Eds.), *Attention and performance IX.* Hillsdale, NJ: Erlbaum, 459-472.

Klaudy, K. (2004). Bevezetés a fordítás elméletébe [Introduction to the Theory of Translation]. Budapest: Scholastica.

Koops, W. (1986). Multiphasic growth curve analysis. *Growth*, *50*, 169-177.

Köpke, B., & Nespoulous, J.-L. (2006). Working memory performance in expert and novice interpreters. *Interpreting 8(1)*, 1-23.

Korpal, P. (2016). *Linguistic and Psychological Indicators of Stress in Simultaneous Interpreting.* (Unpublished doctoral dissertation). Poznan: Adam Mickiewicz University.

Korpal, P. & Stachowiak, K. (2018). The whole picture: processing of numbers and their context in simultaneous interpreting. *Poznań Studies in Contemporary Linguistics, 54*(3), 335-354.

Kramer, J. H., Delis, D. C., Kaplan, E., O'Donnell, L. & Prifitera, A. (1997). Developmental sex differences in verbal learning. *Neuropsychology*, *11*(5), 577-584.

Krick, C., Behrent, S., Reith, W. & Franceschini, R. (2005). Das gläserne Hirn des Dolmetschers:
Vorläufige Forschungsergebnisse über Code-Switching bei mehrsprachigen Personen [The glass brain of interpreters: preliminary research findings on Code-Switching in multilinguals]. *MDÜ-Mitteilungne für Übersetzer und Dolmetscher, 51*(6), 6-9.

Kurz, I. (1993). Conference interpretation: Expectations of different user groups. *The Interpreters' Newsletter, 5*, 13–21.

Kurz, I. (1996). ‘atching the brain at work - An exploratory study of EEG changes during simultaneous interpreting. *The Interpreter's Newsletter, 6*, 3-16.

Kurz, I. (2001). Conference interpreting: quality in the ears of the user. *Meta, 46*(2), 394-409.

Kutas, M. & Hillyard, S. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*(5947), 161-3.

Lakin, J. M. (2013). Sex differences in reasoning abilities: surprising evidence that male–female ratios in the tails of the quantitative reasoning distribution have increased. *Intelligence, 41*, 263–274.

Lamberger-Felber, H. (2001). Text-oriented research into interpreting. Examples from a case study. *Hermes, 26*, 39-63.

Lamberger-Felber, H. (2003). Performance variability among conference interpreters. Examples from a case study. In A. Collados Aís, M. Fernández Sánchez & D. Gile (Eds). *La evaluación de la. calidad en interpretación: investigación [Evaluation of quality in interpreting: investigation*. Granada: Comares, 147- 168.

Lambert, S. (1989). Simultaneous interpreters: One ear may be better than two. *TTR : traduction, terminologie, rédaction, 2(1)*, 153–162.

Lass, N. J., Hughes, K., Bowyer, M., Waters, L. & Bourne, V. (1976). Speaker sex identification from voiced, whispered, and filtered isolated vowels. *Journal of the Acoustical Society of America, 59*, 675-678.

Lass, N. J. & Puffenberger, M. D. (1971). A Comparative study of rate evaluations of experienced and inexperienced listeners. *Quaterly Journal of speech, 57*(1), 89-93.

Lawson, E. A. (1967). Attention and Simultaneous Translation. *Language and Speech, 10(1)*, 29–35.

Lederer, M. (1978). Simultaneous interpretation. Units of meaning and other features. In D. Gerver & W. Sinaiko (Eds.), *Language Interpretation and Communication*. New York: Plenum, 323-332.

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

Lee, T.-H. (2002). Ear Voice Span in English into Korean Simultaneous Interpretation. *Meta, 47*(4), 596–606.

Lehre, A. C., Lehre, K. P., Laake, P. & Danbolt, N. C. (2009). Greater intrasex phenotype variability in males than in females is a fundamental aspect of the gender differences in humans. *Developmental Psychobiology, 51*, 198–206.

Levelt, W. (1983). Monitoring and self-repair in speech. *Cognition, 14*, 41-104.

Levelt, W. (1989). *Speaking: From intention to articulation*. Cambridge: The MIT Press.

Levelt, W. (1999). Models of word production. *Trends in Cognitive Sciences, 3*(6), 223-232.

Lim, H.-O. (2005). Meeting Students' Expectations. *Forum 1*(3), 175-204.

Liontou, K. (2011). Strategies in German-to-Greek simultaneous interpreting: A corpus-based approach. *Gramma, 19*, 37-56.

Lippa, R.A.(2005). *Gender, Nature, and Nurture* (2nd ed.) New jersey, Etats Unis : Lawrence Erlbaum Associates

Liu, M. (2001). *Expertise in simultaneous interpreting: A working memory analysis*. (Unpublished doctoral dissertation), the University of Texas at Austin.

Liu, M. (2008). How do experts interpret? Implications from research in Interpreting Studies and cognitive science. In G. Hansen, A. Chesterman & H. Gerzymisch-Arbogast (Eds.),
Efforts and Models in Interpreting and Translation Research: A tribute to Daniel Gile. Amsterdam: John Benjamins, 159–177.

Liu, M., Schallert, D. & Carroll P. (2004). Working memory and expertise in simultaneous interpreting. *Interpreting 6*(1), 19-42.

Loonstra, A., Tarlow, A. & Sellers A. (2001). COWAT metanorms across age, education and gender. *Applied Neuropsychology*, *8*(3), 161-166.

Lynn, R., Fergusson, D. M., & Horwood, L. J. (2005). Sex differences on the WISC-R in New Zealand. *Personality and Individual Differences, 39*, 103–114.

Lynn, R. & Irwing, P. (2008). Sex differences in mental arithmetic, digit span, and g defined as working memory capacity. *Intelligence, 36*(3), 226-235.

MacGregor, L., Corley M. & Donaldson D. (2009). Not all disfluencies are equal: The effects of disfluent repetitions on language comprehension. *Brain and Language, 111*(1), 36-45.

MacGregor, L., Corley M. & Donaldson D. (2010). Listening to the sound of silence: Investigating the consequences of disfluent silent pauses in speech for listeners. *Neuropsychologia, 48*(14), 3982-3992.

Machin, S. & Pekkarinen, T. (2008). Assessment. Global sex differences in test score variability. *Science, 322*, 1331–1332.

MacKay, D. (1973). Aspects of the theory of comprehension, memory and attention. *The Quarterly journal of experimental psychology, 25*(1), 22-40.

Magnifico, C. & Defrancq, B. (2016). Impoliteness in interpreting: a question of gender? *Translation and Interpreting*, *8*(2), 26-45.

Magnifico, C. & Defrancq, B. (2017). Hedges in conference interpreting: The role of gender. *Interpreting*, *19*(1), 21-46.

Magnifico, C. & Defrancq, B. (Submitted). Self-repair as a norm-related strategy in simultaneous interpreting and its implications for gendered approaches to interpreting. *The Translator.*

Maitland, S., Herlitz, A., Nyberg L., Bäckman L. & Nilsson L.-G. (2004). Selective sex differences in declarative memory. *Memory & Cognition, 32*(7), 1160-1169.

Majeres, R. L. (1999). Sex differences in phonological processes: speeded matching and word reading. *Memory & Cognition, 27*(2), 246-253.

Marwha, D., Halari, M. & Eliot, L. (2016). Meta-analysis reveals a lack of sexual dimorphism in human amygdala volume. *Neuroimage, 147*, 282–294.

Mason, M. (2008). *Courtroom Interpreting*. Maryland: University Press of America.

Mazure, C. M. & Swendsen, J. (2016). Sex differences in Alzheimer's disease and other dementias. *The Lancet Neurology, 15*, 451–452.

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

McCarthy, M. M., & Arnold, A. P. (2011). Reframing sexual differentiation of the brain. N*ature neuroscience, 14(6)*, 677.

Mead, P. (2000). Control of pauses by trainee interpreters in their A and B languages. *The Interpreters' Newsletter, 10*, 89–102.

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

Meiran, N., Hommel, B., Bibi, U. & Lev, I. (2002). Consciousness and control in task switching. *Consciousness and Cognition, 11*, 10-33.

Meuleman, C. & Van Besien, F. (2009). Coping with extreme speech conditions in simultaneous interpreting. *Interpreting, 11*(1), 20-34.

Mellinger, C. D. & Hanson, T. A. (2017). Quantitative Research Methods in Translation and Interpreting Studies. New York, NY: Routledge.

Miller, D. I. & Halpern, D. F. (2014). The new science of cognitive sex differences. Trends in Cognitive Sciences, 18(1), 37-45.

Mizuno, A. (2017). Simultaneous Interpreting and Cognitive Constraints. *Bulletin of the College of Literature*, *58*, 1-28.

Monacelli, C. (2000). A nod's as good as a wink to a blind horse. An analysis of the processes of documentary film interpreting. RM Bollettieri Bosinelli et al (eds), 133-152.

Monacelli, C. (2009). Self-preservetion in simultaneous interpreting : Surviving the role. Benjamins translation library

Monacelli, C. (2018).Gender and social class in TV series scandal : Patterns of repetition, pronominal reference and refraction, *International journal of novel research in humanity and social sciences, 5*(4), 19-29.

Moser, B. (1978) Simultaneous interpretation: A hypothetical model and its practical application. In D. Gerver & W. Sinaiko (Eds.), *Language interpretation and communication*. New York/London: Plenum Press, 353-368.

Moser-Mercer, B. (2010). The search for neuro-physiological correlates of expertise in interpreting, 263-287.

Nooteboom, S. (1969). The tongue slips into patterns. In A. G. Sciarone, A. J.van Essen, A. A. van Raad (Eds.), *Nomen (Society): Leyden studies in linguistics and phonetics*. The Hague: Mouton Publishers, 114-132.

Nord, C. L., Valton, V., Wood, J. & Roiser, J. P. (2017). Power-up: a re-analysis of 'power failure' in neuroscience using mixture modelling. *Journal of Neuroscience, 37*, 8051–8061.

Nordet, L. & Voegtlin, L. (1998). Les tests d'aptitude pour l'interprétation de conférence. La mémoire. [Aptitude testing for conference interpreting. Memory]. Unpublished master thesis. ETI, University of Geneva.

Oléron, P. & Nanpon, H. (1965, 2002). Research into simultaneous translation. In F. Pöchhacker & M. Shlesinger (Eds.), *The interpreting studies reader*. London: Routledge, 69-76.

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

Ooms, E. (2008). *La traduction de chiffres en interprétation simultanée.* [Translating numbers in simultaneous interpreting.] (Unpublished master's thesis). Hogeschool Gent, Ghent, Belgium.

Oviatt, S. (1995). Predicting Spoken Disfluencies during Human-Computer Interaction. *Computer Speech and Language, 9*(l), 19-35.

Padilla, P., Bajo, M. T., Cañas, J. J. & Padilla, F. (1995). Cognitive processes of memory in simultaneous interpretation. In J. Tommola (Ed.). *Topics in Interpreting Research (pp. 61-71)*. Turku: University of Turku Press.

Padilla, F., Bajo, M.T. & Macizo, P. (2005). Articulatory suppression in language interpretation: Working memory capacity, dual tasking and word knowledge.Bilingualism. *Language and Cognition 8(3)*, 207-219.

Pietschnig, J., Penke, L., Wicherts, J.M., Zeiler, M. & Voracek, M. (2015). Meta-analysis of associations between human brain volume and intelligence differences: How strong are they and what do they mean? *Neuroscience & Biobehavioral Reviews, 57*, 411–432.

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

Pio, S. (2003). The relation between ST delivery rate and quality in simultaneous interpretation. *The Interpreters' Newsletter, 12*, 69–100.

Plevoets, K. & Defrancq, B. (2016). The effect of informational load on disfluencies in interpreting. A corpus-based regression analysis. *Translation and Interpreting Studies, 11*(2), 202-224.

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)*. *Interpreting, 20*(1), 1-29.

Pöchhacker's, F. (1994). Simultaneous interpretation: "Cultural Transfer" or "Voice-over text?" Snell-Hornby et al. (eds) 169–178.

Pöchhacker, F. (1995). Slips and Shifts in Simultaneous Interpreting. In J. Tommola (Ed.), *Topics in Interpreting Research*. Turku: The University of Turku, Centre for Translation and Interpreting, 73-90.

Pöchhacker, F. (2015). Models. In F. Pöchhacker (Ed.), *The Routledge Encyclopedia of Interpreting Studies*. New York, NY: Routledge, 263–268.

Pöchhacker, F. & Zwischenberger, C. (2010, March 15). Survey on quality and role: conference interpreters' expectations and self-perceptions. Retrieved from http://aiic.net/p/3405 (accessed 8 October 2018)

Postma, A. (2000). Detection of errors during speech production: a review of speech monitoring models. *Cognition, 77*(2), 97-132.

Ritchie, S. J., Cox, S. R., Shen, X., Lombardo, M. V., Reus, L. M., Alloza, C., Harris, M. A., Alderson, H. L., Hunter, S. Neilson, E.... & Deary, I. J. (2018). Sex differences in the adult human brain: Evidence from 5216 UK Biobank Participants. *Cerebral Cortex, 28* (8), 2959–2975.

Roivainen, E. (2011). Gender differences in processing speed: A review of recent research. *Learning and Individual Differences, 21*(2), 145–149.

Rouvier, M., Dupuy, G., Gay, P., Khoury, E., Merlin, T. & Meignier, S. (2013, August). An Open-source State-of-the-art Toolbox for Broadcast News Diarization. In *Proceedings of Interspeech*. Lyon.

Ruigrok, A. N., Salimi-Khorshidi, G., Lai, M. C., Baron-Cohen, S., Lombardo, M. V., Tait, R. J. & Suckling, J. (2014). A meta-analysis of sex differences in human brain structure. *Neuroscience & Biobehavioral Reviews, 39*, 34–50.

Ruff, R. M. & Light, R. H. & Quayhagen, M. (1989). Selective Reminding Tests: a normative study of verbal learning in adults. *Journal of clinical and experimental neuropsychology, 11,* 539-50.

Rumelhart D.E.& McClelland,J.L.(1986).*Explorations of microstructure of cognition : foundations of human reasoning.*

Russo, M. (2018). Speaking Pattern and Gender in the European Parliament Interpreting Corpus: A Quantitative Study as a Premise for Qualitative Investigations. In M. Russo, C. Bendazzoli & B. Defrancq (Eds.), *Making Way in Corpus-based Interpreting Studies.* Singapore: Springer, 115-132.

Ryan, R. (2015, October 29). Why so few men? : Gender imbalance in conference interpreting. Retrieved from http://aiic.net/p/7347, (accessed 8 October 2018).

Salevsky, H.(1993) . The distinctive nature of interpreting studies. *Target, 5*(2), 149-167.

Schachter, S., Christenfeld, N., Ravina, B. & Bilous, F. (1991). Speech disfluency and the structure of knowledge. *Journal of Personality and Social Psychology, 60*, 362-367.

Schmidt, T. (2012, January). EXMARaLDA and the FOLK tools. In N. Calzolari, K. Choukri, T. Declerck, M. Doğan, B. Maegaard, J. Mariani,... & S. Piperidis (Eds.), *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12).* Istanbul: European Language Resources Association.

Schmidt, T. & Wörner, K. (2009). EXMARaLDA: creating, analyzing and sharing spoken language corpora for pragmatic research. *Pragmatics, 19*(4), 565-582.

Schnack, H.G., Van Haren, N.E., Brouwer, R.M., Evans, A., Durston, S., Boomsma, D.I., Kahn, R.S. & Hulshoff Pol, H.E. (2014). Changes in thickness and surface area of the human cortex and their relationship with intelligence. *Cerebral Cortex, 25*, 1608–1617.

Schuller, B. W. & Batliner, A. M. (2013). Computational Paralinguistics: Emotion, Affect and Personality in Speech and Language Processing. Hoboken: Wiley.

Seeber, K. (2011). Cognitive load in simultaneous interpreting: Existing theories – new models. *Interpreting, 13*(2), 176-204.

Seeber, K. (2012). Multimodal input in simultaneous interpreting: An eye-tracking experiment. In L. Zybatov, A. Petrova & M. Ustaszewski (Eds.), *Proceedings of the 1st International Conference TRANSLATA, Translation & Interpreting Research: yesterday - today – tomorrow.* Forum Translationswissenschaft. Frankfurt a.M.: Peter Lang, 341–347.

Seeber, K.(2013). Cognitive load in simultaneous interpreting: Measures and Methods. *Target : international journal of translation studies, 25*(1), 18-32.

Seeber, K. (2015). Simultaneous Interpreting. In H. Mikkelsen, & R. Jourdenais (Eds.), *The Routledge Handbook of Interpreting*. Oxon/New York: Routledge.

Seleskovitch, D. (1975). *Langage, Langue et Mémoire. Etude de la Prise de Notes en Interpretation Consécutive*. [Language, languages and memory. A study of note-taking in consecutive interpreting.] Paris: Minard Lettres Modernes.

Setton, R. (1999). *Simultaneous interpretation: A cognitive-pragmatic analysis.* Amsterdam/Philadelphia: John Benjamins.

Setton, R. & Dawrant, A. (2016). *Conference Interpreting: A complete course.* Amsterdam/Philadelphia: John Benjamins.

Shallice, T. & Butterworth, B. (1977). Short-term memory impairment and spontaneous speech. *Neuropsychologia, 15*, 729-735.

Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial-ordering mechanism in sentence production. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing.* Hillsdale, NJ: Lawrence Erlbaum, 295–342.

Shewan, C. M. & Henderson, V. L. (1988). Analysis of spontaneous language in the older normal population. *Journal of Communication Disorders, 21*(2), 139-154.

Shlesinger, M. (1998). Corpus-based interpreting studies as an offshoot of corpus-based translation studies. *Meta, 43*(4), 486–493.

Shlesinger, M. (2000). Interpreting as a cognitive process: How can we know what really happens? In S. Tirkkonen-Condit, & R. Jääskeläinen (Eds.), *Tapping and Mapping the Processes of Translation and Interpreting.* Amsterdam/Philadelphia: John Benjamins, 3–15.

Shlesinger, M. (2003). Effects of presentation rate on working memory in simultaneous interpreting. *The Interpreter's Newsletter, 12*, 37–49.

Shriberg, E. (1994). *Preliminaries to a Theory of Speech Disfluencies* (Unpublished doctoral dissertation). University of California, Berkeley, United States of America.

Shriberg, E. (1996). Disfluencies in Switchboard. In *Proceedings of the International Conference on Spoken Language Processing* (Vol. Addendum, 11–14). Philadelphia, PA, United States of America.

Solianik, R., Brazaitis, M. & Skurvydas, A. (2016). Sex-related differences in attention and memory. *Medicina, 52*(6), 372-377.

Sommer, I.E., Aleman, A., Somers, M., Boks,M.P. & Kahn, R.S.(2008). Sex differences in handedness, asymmetry of the planum temporale and functional language lateralization. *Brain Research, 1206*, 76–88.

Spencer, S. J., Steele, C. M. & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35*, 4-28.

Stavrakaki, S., Megari, K., Kosmidis, M. H., Apostolidou, M. & Takou, E. (2012). Working memory and verbal fluency in simultaneous Interpreters. *Journal of Clinical and Experimental Neuropsychology, 34*(6), 624–33.

Sternberg, S. (1969). Memory scanning: Mental processes revealed by reaction-time experiments. *American Scientist, 4*, 421–457.

Swerts, M. (1998). Filled pauses as markers of discourse structure. *Journal of Pragmatics, 30*, 485–496.

Szakos, J. & Glavitsch, U. (2007, August). *SpeechIndexer in action: Managing endangered Formosan languages.* Paper presented at the 8th annual Conference of the International Speech Communication Association. Antwerp, Belgium.

Tan, A., Ma, W., Vira, A., Marwha, D. & Eliot, L. (2016). The human hippocampus is not sexually-dimorphic: meta-analysis of structural MRI volumes. *Neuroimage, 124*, 350–366.

Thurmair, M. (1991). Warten auf das Verb: Die Gedaechtnisrelevanz der Verbklammer im Deutschen [Waiting for the verb: the importance of memory in German verbal parentheses]. *Jahrbuch Deutsch als Fremdsprache, 17*, 174-202.

Timarová, S. (2012). *Working Memory in Simultaneous Interpreting* (Unpublished doctoral dissertation). KULeuven, Antwerp, Belgium.

Timarová, S. (2015). Working Memory. In F. Pöchhacker (Ed.). *The Routledge Encyclopedia of Interpreting Studies*. New York, NY: Routledge, 443–446.

Timarová, S., Čeňková, I., Meylaerts, R., Hertog, E., Szmalec, A., & Duyck, W. (2014). Simultaneous interpreting and working memory executive control. *Interpreting, 16(2)*, 139–168.

Timarová, S., Dragsted, B. & Hansen, I. G. (2011). Time lag in translation and interpreting: A methodological exploration. In C. Alvstad, A. Hild, & E. Tiselius

(Eds.), *Methods and strategies in process research: Integrative approaches in translation studies*. Amsterdam: John Benjamins, 121–146.

Tissi, B. (2000). Silent pauses and disfluencies in simultaneous interpretation: A descriptive analyzes. *The Interpreters' Newsletter, 10*, 103–127.

Tóth, A. (2011). Speech disfluencies in simultaneous interpreting: a mirror on cognitive processes. *SKASE Journal of Translation and Interpretation, 5*(2), 23–31.

Trahan, D. E. & Quintana, J. W. (1990). Analysis of gender effects upon verbal and visual memory performance in adults. *Archives of Clinical Neuropsychology, 5*(4), 325-334.

Treisman, A. (1965). The Effects of Redundancy and Familiarity on Translation and Repeating Back a Native and Foreign Language. *British Journal of Psychology, 56*(4), 369-379.

Vallar, G. & Baddeley, A. (1984). Fractionation of working memory. Neuropsychological evidencefor a phonological short-term store. *Journal of Verbal Learning and Verbal Behaviour, 23*, 151-162.

Van Besien, F. (1999). Anticipation in Simultaneous Interpretation. *Meta, 44*(2), 250–259.

Van der Sluis, S., Posthuma, D., Dolan, C. V., de Geus, E. J., Colom, R., & Boomsma, D. I. (2006). Sex differences on the Dutch WAIS-III. *Intelligence, 34,* 273–289.

Van Haeringen, C. (1947). Tangconstructies' en reacties daarop. *De Nieuwe Taalgids, 40*, 1-7.

Vandenbosch, L. (1985). PpoverV: een funktioneel-pragmatisch principe. *Antwerp Papers in Linguistics, 46*, 1-72.

Vander Beken, A. (2013). *La traduction simultanée de chiffres par des interprètes masculins et féminins: une étude de corpus.* [The simultaneous translation of numbers by male and female interpreters: a corpus-based study.] (Unpublished master's thesis). Hogeschool Gent, Ghent, Belgium.

Vigliocco, G. (1998). Separating hierarchical relations and word order in language production: Is proximityconcord syntactic or linear? *Cognition, 68*, 13-29.

Voor, J. B. & Miller, J. M. (1965). The effect of practice on the comprehension of worded speech. *Speech Monographs, 32*, 452-455.

Watanabe, M., Keikichi H., Yasuharu D. & Nobuaki M. (2008). Filled pauses as cues to the complexity of up-coming phrases for native and non-native listeners. *Speech Communication, 50*(2), 81–94.

Welford, A. T. (1952). The ‘psychological refractory period’ and the timing of high speed performance — a review and a theory. *British Journal of Psychology, 43*, 2–19.

Wierenga, L. M., Sexton, J. A., Laake, P., Giedd, J. N. & Tamnes, C. K. (2017). A key characteristic of sex differences in the developing brain: greater variability in brain structure of boys than girls. *Cerebral Cortex, 28*, 2741–2751.

Willems, A. (2017). *Uit het veld geslagen: een corpusgebaseerd multivariaat onderzoek naar de positie van voorzetselconstituenten in de Nederlandse journalistieke bijzin* [Corpus-based multivariate study on the position of proposition’s components in Dutch journalistic subordinate clauses] (Unpublished doctoral dissertation). Universiteit Gent, Ghent, Belgium.

Yairi, E. & Ambrose N. A. (1992). Longitudinal study of stuttering in children: A preliminary report. *Journal of Speech and Hearing Research, 35*, 755–760.

Yap, T. F. (2012). *Speech production under cognitive load: Effects and classification* (Unpublished doctoral dissertation). University of New South Wales, Sydney, Australia.

Zillmann, D. (1971). Excitation transfer in communication-mediated aggressive behavior. *Journal of Experimental Social Psychology, 7*, 419-434.

Zwart, J. (1990). *PP Extraposition from NP in Dutch.* Unpublished manuscript. Groningen: Rijksuniversiteit Groningen.

Zwart, J. (2011). *The syntax of Dutch.* Cambridge: Cambridge University Press.

Appendix 1

ProcessEXB

Description

J. Collard

CONTENTS

Introduction

Purpose

The Process EXB Tool (ProcessEXB) aims to support the analysis of a set of EXB files to extract data and to compute related statistics. The ProcessEXB works on EXB file items, TIER Events items and Speech Segments items.

It provides different output files, most of them being formatted as CSV files.

A log and a listing of statistical results are also provided

The scope of the ProcessEXB extends to

- The set of EXB files, that constitutes all the data necessary to fulfil its purpose.
- All the functionality necessary to fulfil its purpose, as follows:
    - The Human Machine Interface, supporting the definition of the directory containing the set of EXB files to analyse
    - Some configuration data.
    - The processing of the input files and the creation of the output files.

The ProcessEXB will run on any host supporting TCL Script Language.

Assumptions

This document applies to ProcessEXB version greater or equal to 2.0.0

Associated documents

Title	Reference	Issue
Audio_Processing_Description		

Constraints

User characteristics

The final User is assumed experimented in EXB Files data and having a minimal knowledge in statistics.

The Software Maintainer is assumed experimented in TCL Script Language.

Production constraints

The ProcessEXB was designed for adaptability to :

The evolving requirements for analysis.

The specific requirements from user projects.

Requirements Specification

Interface specifications

List of interfaces

The ProcessEXB input data is a set of EXB Files.

An EXB File is expected containing:

- Attributes applying to the Speech
- Attributes applying to the Speakers
- Time_Data defining the Time Tags
- The Original Speaker Tier (ORIG_TIER) characterized by display-name == 'original'
- The Interpreter Speaker Tier (INTERP_TIER) characterized by display-name == 'interpreter'
- The Annotated Original Speaker Tier (ANNOTATED_ORIG_TIER) characterized by the same speaker as the ORIG_TIER and type == 'a'
- The Annotated Interpreter Tier (ANNOTATED_INTERP_TIER) characterised by the same speaker as the INTERP_TIER and type == 'a'

An EXB File can contain:

1. The Description Original Speaker Tier (DESCRIPTION_ORIG_TIER) characterized by the same speaker as the ORIG_TIER and type == 'd'
2. The Description Interpreter Tier (DESCRIPTION_INTERP_TIER) characterised by the same speaker as the INTERP_TIER and type == 'd'
3. The Link Original Speaker Tier (LINK_ORIG_TIER) characterized by the same speaker as the ORIG_TIER and type == 'l'
4. The Link Interpreter Tier (LINK_INTERP_TIER) characterised by the same speaker as the INTERP_TIER and type == 'l'

The User oriented configuration data are included within the script. They are the followings:

- The PROCESS_PHASE to distinguish the two main parts of the script: “AUDIO” or “EXB” (For “AUDIO” refer to the companion document)
- The Output_Dir_EXB (_AUDIO) to define the script output directory
- The Output_Dir_EXB_ResultsForAudio to indicate where to retrieve results from the AUDIO phase, to include within the EXB phase results (INTERP_TIER__Gender_M, INTERP_TIER__SpeakerID_M)
- The Result_Tag (default 'ALL') to tag every result file.

The variables List_Log_Verbosity and List_Log_Topics controls the logging.

The ProcessEXB output data are the followings

- The log file
- The listing of statistical data
- The CSV formatted files

The CSV files are Text and Excel compatible files. Every line contains data items separated by semicolon. The content of any line is described by lists given in appendices. To avoid duplicating common data items description, a common list is defined, whose name can be 'included' within the description of the other specific lists. Within those lists, a “line” contains 3 items as follows:

- The data title name
- The data program internal name
- An optional formatting instruction

The HMI interface only supports the selection of the input directory.

Functional requirements

Description of functions

The ProcessEXB functions are defined and progressively refined hereafter and within sections starting from 0

At launch, the ProcessEXB shall request the User for the selection of the Directory containing the set of EXB files

For evey EXB file, the ProcessEXB shall proceed to the parsing and related work:

- It shall parse the XML formatted EXB file and convert it into a set of embedded TCL name-value Lists
- It shall resolve the Speech(s) data into lists of Attribute_Name – Attribute_Value
- It shall resolve the Speaker(s) data into lists of Attribute_Name – Attribute_Value
- It shall resolve the time data (TLI) into a list TimeTag – TimeValue
- It shall resolve the lists of events from the different considered TIERs
- It shall collect / compute the data with a unique instance per EXB file and save them

For evey EXB file, the ProcessEXB shall analyse the associated events within the two Annotated Tiers:

- Thanks to the annotations (NNN=*event data*), it shall detect the associated ORIG and INTERP Events, characterize them and compute the different delays (S2S, M2M, E2S, M2S, LOG_S2S: refer to section for details)
- It save the Events related data

As applicable, for evey EXB file, the ProcessEXB shall analyse the associated events within the two Description Tiers:

- Thanks to the annotations (NNN=*event data*), it shall detect the associated ORIG and INTERP Events, compute the NUMBER_DECALAGE_hsec and extract the following data: NUMBER_CATEGORY, NUMBER_RENDITION (refer to section for details)
- It save the Events related data

As applicable, for evey EXB file, the ProcessEXB shall analyse the associated events within the two Link Tiers:

- Thanks to the annotations (NNN=xxxxx), it shall detect the associated ORIG and INTERP Events, compute the LISTOFITEMS_DECALAGE_hsec and extract the following data: ORIGINAL_LISTOFITEMS_ORDER, LISTOFITEMS_NUMBEROFWORDS, LISTOFITEMS_RENDITION (refer to section for details)
- It save the Events related data

Extracting the data from the analysed Tiers

The following processing shall be performed for the different analysed tiers.

From the list of events, the script shall extract

- The FullText as follows:
    - For every event
        - It shall extract the event_text,
        - It filters any “<![CDATA[ ]]>” or “((any chars))” from that event_text,
        - It progressively concatenates the filtered event_text data
    - The concatenated text is filtered to suppress the following patterns: “[any chars]”, “(any chars)”, “space(s))ANY_NOT_SPACE/” or euh or hum or hm
- The ConcatenatedText as follows:
    - For every event
        - It shall extract the event_text,
        - It progressively concatenates the (not filtered) event_text data

For the ORIG_TIER and the INTERP_TIER, the script shall define / compute the following data:

- The FullText_WordsCount: as follows: the count of occurrences of the regular expression pattern '[[:alnum:]_]+ ' (a sequence of alphanumeric or underscore characters)
- The AverageWordLength_char, the average word length in character count
- The UniqueWordsCount
- The Euh_count, the count of pattern '(euh|hum|hm)\[:]?' within the concatenated text
- The ElongatedSoundsCount, the count of (regular expression) pattern "(\[\[:space:]]+\[\[]\[\[:space:]]*\[\[:alpha:]]+\[\[:space:]]*\[:]\[\[:space:]]*\[\[:alpha:]]*\[\]]\[\[:space:]]+)" within the concatenated text (or in other terms, the pattern “some spaces[optional space followed by some alphanumeric characters followed by optional space:optional space followed by optional alphanumeric characters] some spaces => [moete:n])
- The FALSE_STARTCount,the count of pattern "(\[\[:space:]]+\[\[:alpha:]]+\[\/]\[\[:space:]]+)" within the concatenated text
- The SILENT_PAUSESCount, the count of pattern "\[\(]{2,2}\[^)]+\[\)]{2,2}" within the concatenated text. Eg: ((0,4s))
- The TotalDurationOfSILENT_PAUSES_sec by adding the different pause durations
- The AverageSentenceLength_Words, by computing the ratio FullText_WordsCount / (SILENT_PAUSESCount + 1)
- The CountOfNumbers within the related Description_ConcatenatedText
- The CountOfListOfItems, the count of pattern "\[^\[:digit:]\.\,]\[\[:digit:]\.\,]+\[^\[:digit:]\.\,]" within the related Link_ConcatenatedText
- The tier Duration_sec, as Tier_TimeTag_End - Tier_TimeTag_Start

- The WordsPerMin
- The UniqueWordsCountPerMin
- The EUHPerMin
- The ElongatedSoundsPerMin
- The FALSE_STARTPerMin
- The SILENT_PAUSESPerMin
- The AverageDurationOfSILENT_PAUSESPerMin_sec

Defining the data from the ANNOTATED_ORIG_TIER and the ANNOTATED_INTERP_TIER

The following processing shall be performed for the following tiers: ANNOTATED_ORIG_TIER, ANNOTATED_INTERP_TIER

The script shall define / compute the following data:

- The ANNOTATED_ORIG_TIER__NumberOfAnnotations, as the count of events within the ANNOTATED_ORIG_TIER
- The ANNOTATED_ORIG_TIER__NumberOfAnnotationsPerMin
- The ANNOTATED_ORIG_TIER__NumberOfAnnotationsPer_50_InterpretedWords as "50 * ANNOTATED_ORIG_TIER__NumberOfAnnotations / $NTERP_TIER__FullText_WordsCount"
- The ANNOTATED_ORIG_TIER__NumberOfMarkedAnnotations as the count of not empty event codes (in fact, the event annotation marks)
- The ANNOTATED_ORIG_TIER__NumberOfMarkedAnnotationsPer_50_InterpretedWords as "50 * ANNOTATED_ORIG_TIER__NumberOfMarkedAnnotations / INTERP_TIER__FullText_WordsCount"

For every not empty Event_Code within the ANNOTATED_ORIG_TIER,the script

- Shall extract the corresponding events within the ANNOTATED_ORIG_TIER and the ANNOTATED_INTERP_TIER
- Shall extract the (resolved) event time tags: ORIG_start, ORIG_end, INTERP_start and INTERP_end
- Shall compute the following delays (hundredth of sec):
  - S2S_DELAY_hsec, Start to Start: (INTERP_start - ORIG_start) * 100
  - M2M_DELAY_hsec , Mid to Mid:(INTERP_start + INTERP_end)/2 - (ORIG_start + ORIG_end)/2) * 100
  - E2S_DELAY_hsec , End to Start: (INTERP_start - ORIG_end) * 100
  - M2S_DELAY_hsec ,Mid to Start: (INTERP_start - (ORIG_start + ORIG_end)/2) * 100

- LOG_S2S_DELAY_hsec

Defining the data from the DESCRIPTION_ORIG_TIER and the DESCRIPTION_INTERP_TIER

For every not empty Event_Code (event mark code: nnn=*event data*) within the DESCRIPTION_ORIG_TIER,the script

- Shall extract the corresponding events within the DESCRIPTION_ORIG_TIER and the DESCRIPTION_INTERP_TIER
- Shall extract, from the ORIG_Text (text from the DESCRIPTION_ORIG_TIER event), the NUMBER_CATEGORY and the NUMBER_RENDITION by applying the regular expression pattern “(\[^-]+)-(\[^-]+)-(.+)” (or in other words: a..a-b..b-c..c)
- Shall validate the NUMBER_CATEGORY as belonging to the list “N2,N3,D2, D3,C2,C3”
- Shall validate the NUMBER_RENDITION, as formed from the following character sets: “A-Za-z0-9/”
- In case the event from the DESCRIPTION_INTERP_TIER is not empty, the script
  - Shall extract the (resolved) event time tags: ORIG_start and INTERP_start
  - Shall compute the NUMBER_DECALAGE_hsec, as (INTERP_start – ORIG_start) * 100

- Otherwise, in case the NUMBER_RENDITION is not "CO", the script shall logs a WARNING

Defining the data from the LINK_ORIG_TIER and the LINK_INTERP_TIER

For every not empty Event_Code (event mark code: nnn=*event data*) within the LINK_ORIG_TIER,the script

- Shall extract the corresponding events within the LINK_ORIG_TIER and the LINK_INTERP_TIER
- In case the events from the LINK_INTERP_TIER are not empty, the script
    - Shall extract the (resolved) event time tags: ORIG_start and INTERP_start
    - Shall compute the LISTOFITEMS_DECALAGE_hsec as (INTERP_start - ORIG_start) * 100
    - Shall extract, from the ORIG_Text (text from the LINK_ORIG_TIER event), the ORIGINAL_LISTOFITEMS_ORDER, the LISTOFITEMS_NUMBEROFWORDS and the LISTOFITEMS_RENDITION by applying the regular expression pattern “(\[^-]+)-(\[^-]+)-(.+)” (or in other words: a..a-b..b-c..c)
    - Shall validate the ORIGINAL_LISTOFITEMS_ORDER as belonging to the list [list 123 1234 12345 123456 1234567 12345678 123456789]
    - Shall validate the LISTOFITEMS_NUMBEROFWORDS as being an integer number
    - Shall validate the LISTOFITEMS_RENDITION, as formed from the following character sets: “A-Za-z0-9/”

Computing the statistics

The script shall compute statistical data about the list of delay data computed from the set of related events within the annotated tiers.

For every delay data type S2S, M2M, E2S, M2S, LOG_S2S, NUMBER_DECALAGE, LISTOFITEMS_DECALAGE, the script

- Shall retrieve the list of related delay values
- Shall compute the MEAN, the STDDEV, the VAR(iance), the MEDIAN
- Shall compute other statistics to save within the output file 'ProcessEXB_statistics..”

Processing Segments

For each target segment duration (10, 30 sec), the script shall retrieve the list of Event_Codes from the ANNOTATED_ORIG_TIER

For each not empty Event_Code, the script

- Shall extract the corresponding events within the ANNOTATED_ORIG_TIER and the ANNOTATED_INTERP_TIER
- Shall extract the (resolved) event time tags: ORIG_start, ORIG_end, INTERP_start and INTERP_end
- Shall compute the S2S_DELAY_hsec (hundredth of sec) and save it within a list associated with the current segment
- Shall define the (next) segment within the ANNOTATED_ORIG_TIER, whose effective duration is closest to the target duration
- In case a new segment is found, the script
  - Shall compute the Current_Segment_MEAN_S2S_DELAY_hsec and the Current_Segment_STDEV_S2S_DELAY_hsec from the list of delays associated with the current segment
  - Shall define within the different tiers (ORIG_TIER, INTERP_TIER, DESCRIPTION_ORIG_TIER,DESCRIPTION_INTERP_TIER,LINK_ORIG_TIER, LINK_INTERP_TIER) the segment whose start and end time tags are closest to the start and end time tag of the reference segment (found within the ANNOTATED_ORIG_TIER)
  - Shall define / compute the set of segment related variables (refer to the list of segment variables within the appendices) , based on the same procedures as for the whole tiers, but applying to the segment related texts (FullText, ConcatenatedText,...)

Computing the RangeValues

When the different EXB files are processed, the script shall compute the RangeValues associated with different variables.(refer to the lists describing the set of varibles with a RangeValue)

For most of the variables, the RangeValue is computed from the variable value, the variable mean and the variable stddev (standard deviation) according to the following rules:

Criteria	**Result**
Value < max(mean - 1.5 * stdev,0)	1
max(mean - 1.5 * stdev,0) <= Value < max(mean - 0.5 * stdev,0)	2
max(mean - 0.5 * stdev,0) <= Value < mean + 0.5 * stdev	3
mean + 0.5 * stdev <= Value < mean + 1.5 * stdev	4
mean + 1.5 * stdev <= Value	5

For some variables, a dedicated procedure shall apply, as follows:

- SPEECH_TIME (Compute_RangeValue_SPEECH_TIME):

Criteria	**Result**
Default	-1
8*60 min <= Value < = 11*60 min	1
11*60 + 1 min <= Value < = 14*60 min	2
14*60 + 1 min <= Value < = 17*60 min	3
17*60 + 1 min <= Value < = 20*60 min	4
20*60 + 1 min <= Value < = 24*60 min	5

- SPEECH_DELIVERY_TYPE (Compute_RangeValue_SPEECH_DELIVERY_TYPE):

Criteria	**Result**
Default	-1
Value == “impromptu”	2
Value == “mixed”	3
Value == “read”	4

While computing the RangeValues based on the mean and the stddev, the script shall write the detailed results within the output file ProcessEXB_Statistics_...txt

Computing statistics for some data subsets

When the different EXB files are processed, the script shall compute statistics on some subsets.

The subsets are: ALL the EXB files, EXB Files characterized by INTERP_TIER__speaker__sex_value == “m” and EXB files characterized by INTERP_TIER__speaker__sex_value == “f”

The concerned variables are S2S_DELAY_hsec, M2M_DELAY_hsec, E2S_DELAY_hsec, M2S_DELAY_hsec, LOG_S2S_DELAY_hsec, NUMBER_DECALAGE_hsec and LISTOFITEMS_DECALAGE_hsec

The computed statistics are mean, min, max, count, stdev, var and median

The script shall write the detailed results within the output file ProcessEXB_Statistics_...txt

Saving the results

Some results are progressively written within the corresponding file: log, ProcessEXB_Statistics.

However, te output files of type *.csv are written at the end of the whole processing, from the internal Tcl name-value lists and according to the list "Titles_Vars_Format" (reproduced hereafter within the appendices).

The following result csv files are created: ProcessEXB_SUMMARY, ProcessEXB_SEGMENTS_10, ProcessEXB_SEGMENTS_30, ProcessEXB_SUMMARY_EVENTS_DETAILS_FROM_ANNOTATED_ORIG_TIER, ProcessEXB_SUMMARY_EVENTS_DETAILS_FROM_DESCRIPTION_ORIG_TIER, ProcessEXB_SUMMARY_EVENTS_DETAILS_FROM_LINK_ORIG_TIER

The script shall also save results to be used by the audio part within the file ResultsEXBForAudio_InterpGenderAndLang.txt

IMPLEMENTATION DESCRIPTION

The ProcessEXB script is witten in TCL (Tool Command Language) freely available. The used TCL is from ActiveTcl

All the used library packages are included within the distribution

The script contains the main execution loops as well as a lot of procedures.

Common set of variables

The following list identifies the set of common variables to be included within the *.csv output files

```
set List_Titles_Vars_Format_COMMON_SET_OF_VARS_SINGLE_INSTANCE_PER_EXB_FILE {

		{SPEECH DATE}  SPEECH_DATE {}
		{SPEECH TIME}   SPEECH_TIME {}
		{SPEECH TITLE}  SPEECH_TITLE {}
		{SPEECH Delivery_type} SPEECH_DELIVERY_TYPE {}
		{SPEECH Speed} SPEECH_SPEED {}
		{SPEECH Duration} SPEECH_DURATION {}
		 {SPEECH Text_length} SPEECH_TEXT_LENGTH {}
		{SPEECH Complexity}  SPEECH_COMPLEXITY {}
		{SPEECH Topic} SPEECH_TOPIC {}
		{INTERPRETER NAME} INTERP_TIER__speaker__Fullname {}
		{INTERPRETER SEX} INTERP_TIER__speaker__sex_value {}
		{INTERPRETER AGE} INTERP_TIER__speaker__Age {}
		{INTERPRETER LANGUAGE} INTERP_TIER__speaker__Language_0 {}
		{INTERPRETER HOURS WORKED BEFORE}
INTERP_TIER__speaker__Hours_worked_before {}
		{INTERPRETER MINUTES WORKED BEFORE}
INTERP_TIER__speaker__Minutes_worked_before {}
		{INTERPRETER NUMBER SPEAKERS BEFORE}
INTERP_TIER__speaker__Number_speakers_before {}
		{INTERPRETER LEVEL TIREDNESS} INTERP_TIER__speaker__Level_tiredness {}
		{INTERPRETER DISFLUENCIES} INTERP_TIER__speaker__Disfluencies {}
		{INTERPRETER SPEECH DURATION (MIN:SEC)} INTERP_TIER__Duration_sec
{FormatDuration_MS Data}
		{INTERPRETER SPEECH DURATION (SEC)} INTERP_TIER__Duration_sec {format
"%.2f" Data}
		{INTERPRETER SPEECH WORDS COUNT } INTERP_TIER__FullText_WordsCount {}

		{INTERPRETER SPEECH WORDS PER MIN}  INTERP_TIER__WordsPerMin {format
"%.2f" Data}
		{INTERPRETER AVERAGE WORD LENGTH (CHAR)}
INTERP_TIER__AverageWordLength_char {format "%.2f" Data}
		{INTERPRETER UNIQUE WORDS COUNT} INTERP_TIER__UniqueWordsCount {}
		{INTERPRETER UNIQUE WORDS PER MIN}
INTERP_TIER__UniqueWordsCountPerMin {format "%.2f" Data}
```

{INTERPRETER AVERAGE SENTENCE LENGTH (WORDS)} INTERP_TIER__AverageSentenceLength_Words {format "%.2f" Data}
{INTERPRETER SENTENCE COMPLEXITY} {} {}

{INTERPRETER NUMBER OF EUH} INTERP_TIER__EUHCount {}
{INTERPRETER EUH PER MIN} INTERP_TIER__EUHPerMin {format "%.2f" Data}
{INTERPRETER NUMBER OF ELONGATED SOUNDS} INTERP_TIER__ElongatedSoundsCount {}
{INTERPRETER ELONGATED SOUNDS PER MIN} INTERP_TIER__ElongatedSoundsPerMin {format "%.2f" Data}
{INTERPRETER NUMBER OF SILENT_PAUSES} INTERP_TIER__SILENT_PAUSESCount {}
{INTERPRETER NUMBER OF SILENT_PAUSES PER MIN} INTERP_TIER__SILENT_PAUSESPerMin {format "%.2f" Data}
{INTERPRETER TOTAL DURATION OF SILENT_PAUSES} INTERP_TIER__TotalDurationOfSILENT_PAUSES_sec {format "%.2f" Data}
{INTERPRETER AVERAGE DURATION OF SILENT_PAUSES PER MIN} INTERP_TIER__AverageDurationOfSILENT_PAUSESPerMin_sec {format "%.2f" Data}

{INTERPRETER NUMBER OF FALSE_START} INTERP_TIER__FALSE_STARTCount {}
{INTERPRETER NUMBER OF FALSE_START PER MIN} INTERP_TIER__FALSE_STARTPerMin {format "%.2f" Data}
{INTERPRETER EVS} INTERP_TIER__speaker__EVS {}
{INTERPRETER NUMBERS IN INTERPRETATIONS} INTERP_TIER__CountOfNumbers {}
{INTERPRETER LISTS OF ITEMS} INTERP_TIER__CountOfListOfItems {}

{INTERPRETER SEX MACHINE} INTERP_TIER__Gender_M {}
{INTERPRETER ID MACHINE} INTERP_TIER__SpeakerID_M {}

{ORIGINAL SPEAKER NAME} ORIG_TIER__speaker__Fullname {}
{ORIGINAL SPEAKER SEX} ORIG_TIER__speaker__sex_value {}
{ORIGINAL SPEAKER AGE} ORIG_TIER__speaker__Age {}
{ORIGINAL SPEAKER LANGUAGE} ORIG_TIER__speaker__Language_0 {}

{ORIGINAL SPEAKER NATIONALITY} ORIG_TIER__speaker__Nationality {}
{ORIGINAL SPEAKER POLITICAL_FUNCTION} ORIG_TIER__speaker__Political_function {}
{ORIGINAL SPEAKER POLITICAL_GROUP} ORIG_TIER__speaker__Political_group {}

{ORIGINAL SPEAKER DISFLUENCIES} ORIG_TIER__speaker__Disfluencies {}

{ORIGINAL SPEECH DURATION (MIN:SEC)} ORIG_TIER__Duration_sec {FormatDuration_MS Data}

{ORIGINAL SPEECH DURATION (SEC)} ORIG_TIER__Duration_sec {format "%.2f" Data}

{ORIGINAL SPEECH WORDS COUNT} ORIG_TIER__FullText_WordsCount {}

{ORIGINAL SPEECH WORDS PER MIN} ORIG_TIER__WordsPerMin {format "%.2f" Data}

{ORIGINAL AVERAGE WORD LENGTH (CHAR)} ORIG_TIER__AverageWordLength_char {format "%.2f" Data}

{ORIGINAL UNIQUE WORDS COUNT} ORIG_TIER__UniqueWordsCount {}

{ORIGINAL UNIQUE WORDS PER MIN} ORIG_TIER__UniqueWordsCountPerMin {format "%.2f" Data}

{ORIGINAL SPEAKER NUMBER OF EUH} ORIG_TIER__EUHCount {}

{ORIGINAL SPEAKER EUH PER MIN} ORIG_TIER__EUHPerMin {format "%.2f" Data}

{ORIGINAL NUMBER OF ELONGATED SOUNDS} ORIG_TIER__ElongatedSoundsCount {}

{ORIGINAL ELONGATED SOUNDS PER MIN} ORIG_TIER__ElongatedSoundsPerMin {format "%.2f" Data}

{ORIGINAL SPEAKER NUMBER OF SILENT_PAUSES} ORIG_TIER__SILENT_PAUSESCount {}

{ORIGINAL SPEAKER NUMBER OF SILENT_PAUSES PER MIN} ORIG_TIER__SILENT_PAUSESPerMin {format "%.2f" Data}

{ORIGINAL SPEAKER TOTAL DURATION OF SILENT_PAUSES} ORIG_TIER__TotalDurationOfSILENT_PAUSES_sec {format "%.2f" Data}

{ORIGINAL SPEAKER AVERAGE DURATION OF SILENT_PAUSES PER MIN} ORIG_TIER__AverageDurationOfSILENT_PAUSESPerMin_sec {format "%.2f" Data}

{ORIGINAL SPEAKER NUMBER OF FALSE_START} ORIG_TIER__FALSE_STARTCount {}

{ORIGINAL SPEAKER NUMBER OF FALSE_START PER MIN} ORIG_TIER__FALSE_STARTPerMin {format "%.2f" Data}

{ORIGINAL SPEECH NUMBERS} ORIG_TIER__CountOfNumbers {}

{ORIGINAL SPEECH LISTS OF ITEMS} ORIG_TIER__CountOfListOfItems {}

{ORIGINAL AVERAGE SENTENCE LENGTH (WORDS)} ORIG_TIER__AverageSentenceLength_Words {format "%.2f" Data}

{ORIGINAL SENTENCE COMPLEXITY} {} {}

{NUMBER OF ANNOTATIONS} ANNOTATED_ORIG_TIER__NumberOfAnnotations {}

{NUMBER OF MARKED ANNOTATIONS} ANNOTATED_ORIG_TIER__NumberOfMarkedAnnotations {}

{NUMBER OF ANNOTATIONS PER MIN} ANNOTATED_ORIG_TIER__NumberOfAnnotationsPerMin {format "%.2f" Data}

{NUMBER OF ANNOTATIONS PER 50 WORDS INTERPRETED} ANNOTATED_ORIG_TIER__NumberOfAnnotationsPer_50_InterpretedWords {format "%.2f" Data}

{NUMBER OF MARKED ANNOTATIONS PER 50 WORDS INTERPRETED} ANNOTATED_ORIG_TIER__NumberOfMarkedAnnotationsPer_50_InterpretedWords {format "%.2f" Data}

{MIN S2S DELAY (HUNDRETH SEC)} MIN_S2S_DELAY_hsec {format "%.2f" Data}

{MAX S2S DELAY (HUNDRETH SEC)} MAX_S2S_DELAY_hsec {format "%.2f" Data}

{MEAN S2S DELAY (HUNDRETH SEC)} MEAN_S2S_DELAY_hsec {format "%.2f" Data}

{STDEV S2S DELAY (HUNDRETH SEC)} STDEV_S2S_DELAY_hsec {format "%.2f" Data}

{MEDIAN S2S DELAY (HUNDRETH SEC)} MEDIAN_S2S_DELAY_hsec {format "%.2f" Data}

{MIN NUMBER DECALAGE (HUNDRETH SEC)} MIN_NUMBER_DECALAGE_hsec {format "%.2f" Data}

{MAX NUMBER DECALAGE (HUNDRETH SEC)} MAX_NUMBER_DECALAGE_hsec {format "%.2f" Data}

{MEAN NUMBER DECALAGE (HUNDRETH SEC)} MEAN_NUMBER_DECALAGE_hsec {format "%.2f" Data}

{STDEV NUMBER DECALAGE (HUNDRETH SEC)} STDEV_NUMBER_DECALAGE_hsec {format "%.2f" Data}

{MEDIAN NUMBER DECALAGE (HUNDRETH SEC)} MEDIAN_NUMBER_DECALAGE_hsec {format "%.2f" Data}

{MIN LISTOFITEMS DECALAGE (HUNDRETH SEC)} MIN_LISTOFITEMS_DECALAGE_hsec {format "%.2f" Data}

{MAX LISTOFITEMS DECALAGE (HUNDRETH SEC)} MAX_LISTOFITEMS_DECALAGE_hsec {format "%.2f" Data}

{MEAN LISTOFITEMS DECALAGE (HUNDRETH SEC)} MEAN_LISTOFITEMS_DECALAGE_hsec {format "%.2f" Data}

{STDEV LISTOFITEMS DECALAGE (HUNDRETH SEC)} STDEV_LISTOFITEMS_DECALAGE_hsec {format "%.2f" Data}

{MEDIAN LISTOFITEMS DECALAGE (HUNDRETH SEC)} MEDIAN_LISTOFITEMS_DECALAGE_hsec {format "%.2f" Data}

{SPEECH TIME RangeValue} SPEECH_TIME_RangeValue {}
{SPEECH Delivery_type RangeValue} SPEECH_DELIVERY_TYPE_RangeValue {}

{INTERPRETER SPEECH DURATION (SEC) RangeValue} INTERP_TIER__Duration_sec_RangeValue {}

{INTERPRETER SPEECH WORDS COUNT RangeValue} INTERP_TIER__FullText_WordsCount_RangeValue {}
{INTERPRETER SPEECH WORDS PER MIN RangeValue } INTERP_TIER__WordsPerMin_RangeValue {}
{INTERPRETER AVERAGE WORD LENGTH (CHAR) RangeValue } INTERP_TIER__AverageWordLength_char_RangeValue {}
{INTERPRETER UNIQUE WORDS COUNT RangeValue} INTERP_TIER__UniqueWordsCount_RangeValue {}
{INTERPRETER UNIQUE WORDS PER MIN RangeValue} INTERP_TIER__UniqueWordsCountPerMin_RangeValue {}

{INTERPRETER AVERAGE SENTENCE LENGTH (WORDS) RangeValue} INTERP_TIER__AverageSentenceLength_Words_RangeValue {}

{INTERPRETER NUMBER OF EUH RangeValue} INTERP_TIER__EUHCount_RangeValue {}
{INTERPRETER EUH PER MIN RangeValue} INTERP_TIER__EUHPerMin_RangeValue {}

{INTERPRETER NUMBER OF ELONGATED SOUNDS RangeValue} INTERP_TIER__ElongatedSoundsCount_RangeValue {}
{INTERPRETER ELONGATED SOUNDS PER MIN RangeValue} INTERP_TIER__ElongatedSoundsPerMin_RangeValue {}

{INTERPRETER NUMBER OF SILENT_PAUSES RangeValue} INTERP_TIER__SILENT_PAUSESCount_RangeValue {}
{INTERPRETER NUMBER OF SILENT_PAUSES PER MIN RangeValue} INTERP_TIER__SILENT_PAUSESPerMin_RangeValue {}
{INTERPRETER TOTAL DURATION OF SILENT_PAUSES RangeValue} INTERP_TIER__TotalDurationOfSILENT_PAUSES_sec_RangeValue {}

{INTERPRETER AVERAGE DURATION OF SILENT_PAUSES PER MIN RangeValue} INTERP_TIER__AverageDurationOfSILENT_PAUSESPerMin_sec_RangeValue {}

{INTERPRETER NUMBER OF FALSE_START RangeValue} INTERP_TIER__FALSE_STARTCount_RangeValue {}
{INTERPRETER NUMBER OF FALSE_START PER MIN RangeValue} INTERP_TIER__FALSE_STARTPerMin_RangeValue {}

{INTERPRETER NUMBERS IN INTERPRETATIONS RangeValue} INTERP_TIER__CountOfNumbers_RangeValue {}
{INTERPRETER LISTS OF ITEMS RangeValue} INTERP_TIER__CountOfListOfItems_RangeValue {}

{ORIGINAL SPEECH DURATION (SEC) RangeValue} ORIG_TIER__Duration_sec_RangeValue {}

{ORIGINAL SPEECH WORDS COUNT RangeValue} ORIG_TIER__FullText_WordsCount_RangeValue {}
{ORIGINAL SPEECH WORDS PER MIN RangeValue} ORIG_TIER__WordsPerMin_RangeValue {}
{ORIGINAL AVERAGE WORD LENGTH (CHAR) RangeValue} ORIG_TIER__AverageWordLength_char_RangeValue {}

{ORIGINAL UNIQUE WORDS COUNT RangeValue} ORIG_TIER__UniqueWordsCount_RangeValue {}
{ORIGINAL UNIQUE WORDS PER MIN RangeValue} ORIG_TIER__UniqueWordsCountPerMin_RangeValue {}

{ORIGINAL SPEAKER NUMBER OF EUH RangeValue} ORIG_TIER__EUHCount_RangeValue {}
{ORIGINAL SPEAKER EUH PER MIN RangeValue} ORIG_TIER__EUHPerMin_RangeValue {}

{ORIGINAL NUMBER OF ELONGATED SOUNDS RangeValue} ORIG_TIER__ElongatedSoundsCount_RangeValue {}
{ORIGINAL ELONGATED SOUNDS PER MIN RangeValue} ORIG_TIER__ElongatedSoundsPerMin_RangeValue {}

{ORIGINAL SPEAKER NUMBER OF SILENT_PAUSES RangeValue} ORIG_TIER__SILENT_PAUSESCount_RangeValue {}

{ORIGINAL SPEAKER NUMBER OF SILENT_PAUSES PER MIN RangeValue} ORIG_TIER__SILENT_PAUSESPerMin_RangeValue {}

{ORIGINAL SPEAKER TOTAL DURATION OF SILENT_PAUSES RangeValue} ORIG_TIER__TotalDurationOfSILENT_PAUSES_sec_RangeValue {}

{ORIGINAL SPEAKER AVERAGE DURATION OF SILENT_PAUSES PER MIN RangeValue} ORIG_TIER__AverageDurationOfSILENT_PAUSESPerMin_sec_RangeValue {}

{ORIGINAL SPEAKER NUMBER OF FALSE_START RangeValue} ORIG_TIER__FALSE_STARTCount_RangeValue {}

{ORIGINAL SPEAKER NUMBER OF FALSE_START PER MIN RangeValue} ORIG_TIER__FALSE_STARTPerMin_RangeValue {}

{ORIGINAL SPEECH NUMBERS RangeValue} ORIG_TIER__CountOfNumbers_RangeValue {}

{ORIGINAL SPEECH LISTS OF ITEMS RangeValue} ORIG_TIER__CountOfListOfItems_RangeValue {}

{ORIGINAL AVERAGE SENTENCE LENGTH (WORDS) RangeValue} ORIG_TIER__AverageSentenceLength_Words_RangeValue {}

{NUMBER OF ANNOTATIONS RangeValue} ANNOTATED_ORIG_TIER__NumberOfAnnotations_RangeValue {}

{NUMBER OF MARKED ANNOTATIONS RangeValue} ANNOTATED_ORIG_TIER__NumberOfMarkedAnnotations_RangeValue {}

{NUMBER OF ANNOTATIONS PER MIN RangeValue} ANNOTATED_ORIG_TIER__NumberOfAnnotationsPerMin_RangeValue {}

{NUMBER OF ANNOTATIONS PER 50 WORDS INTERPRETED RangeValue} ANNOTATED_ORIG_TIER__NumberOfAnnotationsPer_50_InterpretedWords_RangeValue {}

{NUMBER OF MARKED ANNOTATIONS PER 50 WORDS INTERPRETED RangeValue} ANNOTATED_ORIG_TIER__NumberOfMarkedAnnotationsPer_50_InterpretedWords_RangeValue {}

{MEAN S2S DELAY (HUNDRETH SEC) RangeValue} MEAN_S2S_DELAY_hsec_RangeValue {}

{STDEV S2S DELAY (HUNDRETH SEC) RangeValue} STDEV_S2S_DELAY_hsec_RangeValue {}

{MEAN NUMBER DECALAGE (HUNDRETH SEC) RangeValue} MEAN_NUMBER_DECALAGE_hsec_RangeValue {}
{STDEV NUMBER DECALAGE (HUNDRETH SEC) RangeValue} STDEV_NUMBER_DECALAGE_hsec_RangeValue {}

{MEAN LISTOFITEMS DECALAGE (HUNDRETH SEC) RangeValue} MEAN_LISTOFITEMS_DECALAGE_hsec_RangeValue {}
{STDEV LISTOFITEMS DECALAGE (HUNDRETH SEC) RangeValue} STDEV_LISTOFITEMS_DECALAGE_hsec_RangeValue {}

ProcessEXB_SUMMARY

The following list identifies the set of variables to be included within the *.csv output file ProcessEXB_SUMMARY

set List_Titles_Vars_Format_SUMMARY_NOT_RESOLVED {

%INCLUDE% List_Titles_Vars_Format_COMMON_SET_OF_VARS_SINGLE_INSTANCE_PER_EXB_FILE

{IS S2S NORMALLY DISTRIBUTED (.80) ?} ISNORMAL_C80_S2S_DELAY_hsec {}
{IS S2S NORMALLY DISTRIBUTED (.99) ?} ISNORMAL_C99_S2S_DELAY_hsec {}
{GOODNESS OF FIT TO A NORMAL DISTR. FOR S2S } GOODNESS_OF_FIT_TO_NORMAL_S2S_DELAY_hsec {format "%.2f" Data}

{MIN M2M DELAY (HUNDRETH SEC)} MIN_M2M_DELAY_hsec {format "%.2f" Data}
{MAX M2M DELAY (HUNDRETH SEC)} MAX_M2M_DELAY_hsec {format "%.2f" Data}
{MEAN M2M DELAY (HUNDRETH SEC)} MEAN_M2M_DELAY_hsec {format "%.2f" Data}
{STDEV M2M DELAY (HUNDRETH SEC)} STDEV_M2M_DELAY_hsec {format "%.2f" Data}
{MEDIAN M2M DELAY (HUNDRETH SEC)} MEDIAN_M2M_DELAY_hsec {format "%.2f" Data}
{IS M2M NORMALLY DISTRIBUTED (.80) ?} ISNORMAL_C80_M2M_DELAY_hsec {}
{IS M2M NORMALLY DISTRIBUTED (.99) ?} ISNORMAL_C99_M2M_DELAY_hsec {}
{GOODNESS OF FIT TO A NORMAL DISTR. FOR M2M} GOODNESS_OF_FIT_TO_NORMAL_M2M_DELAY_hsec {format "%.2f" Data}

{MIN E2S DELAY (HUNDRETH SEC)} MIN_E2S_DELAY_hsec {format "%.2f" Data}
{MAX E2S DELAY (HUNDRETH SEC)} MAX_E2S_DELAY_hsec {format "%.2f" Data}
{MEAN E2S DELAY (HUNDRETH SEC)} MEAN_E2S_DELAY_hsec {format "%.2f" Data}
{STDEV E2S DELAY (HUNDRETH SEC)} STDEV_E2S_DELAY_hsec {format "%.2f" Data}
{MEDIAN E2S DELAY (HUNDRETH SEC)} MEDIAN_E2S_DELAY_hsec {format "%.2f" Data}
{IS E2S NORMALLY DISTRIBUTED (.80) ?} ISNORMAL_C80_E2S_DELAY_hsec {}
{IS E2S NORMALLY DISTRIBUTED (.99) ?} ISNORMAL_C99_E2S_DELAY_hsec {}
{GOODNESS OF FIT TO A NORMAL DISTR. FOR E2S} GOODNESS_OF_FIT_TO_NORMAL_E2S_DELAY_hsec {format "%.2f" Data}

{MIN M2S DELAY (HUNDRETH SEC)} MIN_M2S_DELAY_hsec {format "%.2f" Data}
{MAX M2S DELAY (HUNDRETH SEC)} MAX_M2S_DELAY_hsec {format "%.2f" Data}
{MEAN M2S DELAY (HUNDRETH SEC)} MEAN_M2S_DELAY_hsec {format "%.2f" Data}
{STDEV M2S DELAY (HUNDRETH SEC)} STDEV_M2S_DELAY_hsec {format "%.2f" Data}
{MEDIAN M2S DELAY (HUNDRETH SEC)} MEDIAN_M2S_DELAY_hsec {format "%.2f" Data}
{IS M2S NORMALLY DISTRIBUTED (.80) ?} ISNORMAL_C80_M2S_DELAY_hsec {}
{IS M2S NORMALLY DISTRIBUTED (.99) ?} ISNORMAL_C99_M2S_DELAY_hsec {}
{GOODNESS OF FIT TO A NORMAL DISTR. FOR M2S} GOODNESS_OF_FIT_TO_NORMAL_M2S_DELAY_hsec {format "%.2f" Data}

{MIN LOG S2S DELAY (HUNDRETH SEC)} MIN_LOG_S2S_DELAY_hsec {format "%.2f" Data}
{MAX LOG S2S DELAY (HUNDRETH SEC)} MAX_LOG_S2S_DELAY_hsec {format "%.2f" Data}
{MEAN LOG S2S DELAY (HUNDRETH SEC)} MEAN_LOG_S2S_DELAY_hsec {format "%.2f" Data}
{STDEV LOG S2S DELAY (HUNDRETH SEC)} STDEV_LOG_S2S_DELAY_hsec {format "%.2f" Data}
{MEDIAN LOG S2S DELAY (HUNDRETH SEC)} MEDIAN_LOG_S2S_DELAY_hsec {format "%.2f" Data}
{IS LOG S2S NORMALLY DISTRIBUTED (.80) ?} ISNORMAL_C80_LOG_S2S_DELAY_hsec {}
{IS LOG S2S NORMALLY DISTRIBUTED (.99) ?} ISNORMAL_C99_LOG_S2S_DELAY_hsec {}
{GOODNESS OF FIT TO A NORMAL DISTR. FOR LOG S2S } GOODNESS_OF_FIT_TO_NORMAL_LOG_S2S_DELAY_hsec {format "%.2f" Data}

```
		{MEAN M2M DELAY (HUNDRETH SEC) RangeValue} MEAN_M2M_DELAY_hsec_RangeValue {}
		{STDEV M2M DELAY (HUNDRETH SEC) RangeValue} STDEV_M2M_DELAY_hsec_RangeValue {}

		{MEAN E2S DELAY (HUNDRETH SEC) RangeValue} MEAN_E2S_DELAY_hsec_RangeValue {}
		{STDEV E2S DELAY (HUNDRETH SEC) RangeValue} STDEV_E2S_DELAY_hsec_RangeValue {}



		{MEAN M2S DELAY (HUNDRETH SEC) RangeValue} MEAN_M2S_DELAY_hsec_RangeValue {}
		{STDEV M2S DELAY (HUNDRETH SEC) RangeValue} STDEV_M2S_DELAY_hsec_RangeValue {}
	}
```

ProcessEXB_SUMMARY_EVENTS_DETAILS_FROM_ANNOTATED_ORIG_TIER

The following list identifies the set of variables to be included within the *.csv output file ProcessEXB_ SUMMARY_EVENTS_DETAILS_FROM_ANNOTATED_ORIG_TIER

```
set List_Titles_Vars_Format_SUMMARY_EVENTS_DETAILS_FROM_ANNOTATED_ORIG_TIER_NOT_RESOLVED {

		%INCLUDE% List_Titles_Vars_Format_COMMON_SET_OF_VARS_SINGLE_INSTANCE_PER_EXB_FILE


		{ANNOTATION TEXT ORIGINAL} ANNOTATED_ORIG_TIER_EVENT__EText {}
		{ANNOTATION TEXT INTERPRETER} ANNOTATED_INTERP_TIER_EVENT__EText {}
		{ANNOTATION TIMETAG START ORIGINAL} ANNOTATED_ORIG_TIER_EVENT__start {format "%.2f" Data}
		{ANNOTATION TIMETAG END ORIGINAL} ANNOTATED_ORIG_TIER_EVENT__end {format "%.2f" Data}

		{ANNOTATION TIMETAG START INTERPRETER} ANNOTATED_INTERP_TIER_EVENT__start {format "%.2f" Data}
		{ANNOTATION TIMETAG END INTERPRETER} ANNOTATED_INTERP_TIER_EVENT__end {format "%.2f" Data}

		{S2S DELAY (HUNDRETH SEC)} S2S_DELAY_hsec {FormatDuration_MS_hsec Data}
		{S2S DELAY (HUNDRETH SEC)} S2S_DELAY_hsec {format "%.2f" Data}
```

```
        {M2M DELAY (HUNDRETH SEC)} M2M_DELAY_hsec {FormatDuration_MS_hsec Data}
        {M2M DELAY (HUNDRETH SEC)} M2M_DELAY_hsec  {format "%.2f" Data}

        {E2S DELAY (HUNDRETH SEC)} E2S_DELAY_hsec {FormatDuration_MS_hsec Data}
        {E2S DELAY (HUNDRETH SEC)} E2S_DELAY_hsec  {format "%.2f" Data}

        {M2S DELAY (HUNDRETH SEC)} M2S_DELAY_hsec {FormatDuration_MS_hsec Data}
        {M2S DELAY (HUNDRETH SEC)} M2S_DELAY_hsec  {format "%.2f" Data}

        {LOG S2S DELAY (HUNDRETH SEC)} LOG_S2S_DELAY_hsec {format "%.4f" Data}

        {MARKER CODE} ANNOTATED_ORIG_TIER_EVENT__ECode {}

        }
```

ProcessEXB_SUMMARY_EVENTS_DETAILS_FROM_DESCRIPTION_ORIG_TIER

The following list identifies the set of variables to be included within the *.csv output file ProcessEXB_ SUMMARY_EVENTS_DETAILS_FROM_DESCRIPTION_ORIG_TIER

```
set List_Titles_Vars_Format_SUMMARY_EVENTS_DETAILS_FROM_DESCRIPTION_ORIG_TIER_NOT_RESOLVED {

        %INCLUDE% List_Titles_Vars_Format_COMMON_SET_OF_VARS_SINGLE_INSTANCE_PER_EXB_FILE


        {DESCRIPTION MARKER CODE} DESCRIPTION_ORIG_TIER_EVENT__ECode {}


        {NUMBER CATEGORY} NUMBER_CATEGORY {}
        {NUMBER RENDITION} NUMBER_RENDITION {}

        {NUMBER DECALAGE  (HUNDRETH SEC)} NUMBER_DECALAGE_hsec {FormatDuration_MS_hsec Data}
        {NUMBER DECALAGE (HUNDRETH SEC)} NUMBER_DECALAGE_hsec {format "%.2f" Data
```

ProcessEXB_SUMMARY_EVENTS_DETAILS_FROM_LINK_ORIG_TIER

The following list identifies the set of variables to be included within the *.csv output file ProcessEXB_ SUMMARY_EVENTS_DETAILS_FROM_LINK_ORIG_TIER

```
set List_Titles_Vars_Format_SUMMARY_EVENTS_DETAILS_FROM_LINK_ORIG_TIER_NOT_RESOLVED {

        %INCLUDE% List_Titles_Vars_Format_COMMON_SET_OF_VARS_SINGLE_INSTANCE_PER_EXB_FILE
```

{LINK MARKER CODE} LINK_ORIG_TIER_EVENT__ECode {}

{ORIGINAL LIST OF ITEMS ORDER} ORIGINAL_LISTOFITEMS_ORDER {}
{LIST OF ITEMS NUMBER OF WORDS} LISTOFITEMS_NUMBEROFWORDS {}
{LIST OF ITEMS RENDITION} LISTOFITEMS_RENDITION {}

{LISTOFITEMS DECALAGE (HUNDRETH SEC)} LISTOFITEMS_DECALAGE_hsec {FormatDuration_MS_hsec Data}

{LISTOFITEMS DECALAGE (HUNDRETH SEC)} LISTOFITEMS_DECALAGE_hsec {format "%.2f" Data}

}

ProcessEXB_SEGMENTS

The following list identifies the set of variables to be included within the *.csv output file(s) ProcessEXB_ SEGMENTS

set List_Titles_Vars_Format_SEGMENTS_NOT_RESOLVED {

%INCLUDE% List_Titles_Vars_Format_COMMON_SET_OF_VARS_SINGLE_INSTANCE_PER_EXB_FILE

{SEGMENT ANNOTATED ORIG TIER Duration(sec)} SEGMENT_ANNOTATED_ORIG_TIER__Duration_sec {format "%.2f" Data}
{SEGMENT ORIG TIER Duration(sec)} SEGMENT_ORIG_TIER__Duration_sec {format "%.2f" Data}
{SEGMENT DESCRIPTION ORIG TIER Duration(sec)} SEGMENT_DESCRIPTION_ORIG_TIER__Duration_sec {format "%.2f" Data}
{SEGMENT INTERP TIER Duration(sec)} SEGMENT_INTERP_TIER__Duration_sec {format "%.2f" Data}
{SEGMENT DESCRIPTION INTERP TIER Duration(sec)} SEGMENT_DESCRIPTION_INTERP_TIER__Duration_sec {format "%.2f" Data}
{SEGMENT ANNOTATED ORIG TIER Start} SEGMENT_AOT_Start {format "%.2f" Data}
{SEGMENT ANNOTATED ORIG TIER End} SEGMENT_AOT_End {format "%.2f" Data}
{SEGMENT ECode Start} SEGMENT_AOT_ECode_Start {}
{SEGMENT ECode End} SEGMENT_AOT_ECode_End {}
{SEGMENT MEAN S2S DELAY hsec} SEGMENT_MEAN_S2S_DELAY_hsec {format "%.2f" Data}
{SEGMENT STDEV S2S DELAY hsec} SEGMENT_STDEV_S2S_DELAY_hsec {format "%.2f" Data}

{SEGMENT INTERPRETER SPEECH WORDS COUNT } SEGMENT_INTERP_TIER__FullText_WordsCount {}

{SEGMENT INTERPRETER SPEECH WORDS PER MIN} SEGMENT_INTERP_TIER__WordsPerMin {format "%.2f" Data}

{SEGMENT INTERPRETER AVERAGE WORD LENGTH (CHAR)} SEGMENT_INTERP_TIER__AverageWordLength_char {format "%.2f" Data}

{SEGMENT INTERPRETER UNIQUE WORDS COUNT} SEGMENT_INTERP_TIER__UniqueWordsCount {}
{SEGMENT INTERPRETER UNIQUE WORDS PER MIN} SEGMENT_INTERP_TIER__UniqueWordsCountPerMin {format "%.2f" Data}

{SEGMENT INTERPRETER AVERAGE SENTENCE LENGTH (WORDS)} SEGMENT_INTERP_TIER__AverageSentenceLength_Words {format "%.2f" Data}

{SEGMENT INTERPRETER NUMBER OF EUH} SEGMENT_INTERP_TIER__EUHCount {}
{SEGMENT INTERPRETER EUH PER MIN} SEGMENT_INTERP_TIER__EUHPerMin {format "%.2f" Data}

{SEGMENT INTERPRETER NUMBER OF ELONGATED SOUNDS} SEGMENT_INTERP_TIER__ElongatedSoundsCount {}
{SEGMENT INTERPRETER ELONGATED SOUNDS PER MIN} SEGMENT_INTERP_TIER__ElongatedSoundsPerMin {format "%.2f" Data}

{SEGMENT INTERPRETER NUMBER OF SILENT_PAUSES} SEGMENT_INTERP_TIER__SILENT_PAUSESCount {}
{SEGMENT INTERPRETER NUMBER OF SILENT_PAUSES PER MIN} SEGMENT_INTERP_TIER__SILENT_PAUSESPerMin {format "%.2f" Data}
{SEGMENT INTERPRETER TOTAL DURATION OF SILENT_PAUSES} SEGMENT_INTERP_TIER__TotalDurationOfSILENT_PAUSES_sec {format "%.2f" Data}
{SEGMENT INTERPRETER AVERAGE DURATION OF SILENT_PAUSES PER MIN} SEGMENT_INTERP_TIER__AverageDurationOfSILENT_PAUSESPerMin_sec {format "%.2f" Data}

{SEGMENT INTERPRETER NUMBER OF FALSE_START} SEGMENT_INTERP_TIER__FALSE_STARTCount {}
{SEGMENT INTERPRETER NUMBER OF FALSE_START PER MIN} SEGMENT_INTERP_TIER__FALSE_STARTPerMin {format "%.2f" Data}

{SEGMENT INTERPRETER NUMBERS IN INTERPRETATIONS} SEGMENT_INTERP_TIER__CountOfNumbers {}

{SEGMENT INTERPRETER SPEECH WORDS PER THEORETICAL SEGMENT DURATION} SEGMENT_INTERP_TIER__WordsPerTheoreticalSegmentDuration {format "%.2f" Data}
{SEGMENT INTERPRETER UNIQUE WORDS PER THEORETICAL SEGMENT DURATION} SEGMENT_INTERP_TIER__UniqueWordsCountPerTheoreticalSegmentDuration {format "%.2f" Data}
{SEGMENT INTERPRETER EUH PER THEORETICAL SEGMENT DURATION} SEGMENT_INTERP_TIER__EUHPerTheoreticalSegmentDuration {format "%.2f" Data}
{SEGMENT INTERPRETER NUMBER OF SILENT_PAUSES PER THEORETICAL SEGMENT DURATION} SEGMENT_INTERP_TIER__SILENT_PAUSESPerTheoreticalSegmentDuration {format "%.2f" Data}
{SEGMENT INTERPRETER AVERAGE DURATION OF SILENT_PAUSES PER THEORETICAL SEGMENT DURATION} SEGMENT_INTERP_TIER__AverageDurationOfSILENT_PAUSESPerTheoreticalSegmentDuration_sec {format "%.2f" Data}
{SEGMENT INTERPRETER NUMBER OF FALSE_START PER THEORETICAL SEGMENT DURATION} SEGMENT_INTERP_TIER__FALSE_STARTPerTheoreticalSegmentDuration {format "%.2f" Data}

{SEGMENT INTERPRETER ELONGATED SOUNDS PER THEORETICAL SEGMENT DURATION} SEGMENT_INTERP_TIER__ElongatedSoundsPerTheoreticalSegmentDuration {format "%.2f" Data}

{SEGMENT ORIGINAL SPEECH WORDS COUNT} SEGMENT_ORIG_TIER__FullText_WordsCount {}
{SEGMENT ORIGINAL SPEECH WORDS PER MIN} SEGMENT_ORIG_TIER__WordsPerMin {format "%.2f" Data}
{SEGMENT ORIGINAL AVERAGE WORD LENGTH (CHAR)} SEGMENT_ORIG_TIER__AverageWordLength_char {format "%.2f" Data}

{SEGMENT ORIGINAL UNIQUE WORDS COUNT} SEGMENT_ORIG_TIER__UniqueWordsCount {}
{SEGMENT ORIGINAL UNIQUE WORDS PER MIN} SEGMENT_ORIG_TIER__UniqueWordsCountPerMin {format "%.2f" Data}

{SEGMENT ORIGINAL SPEAKER NUMBER OF EUH} SEGMENT_ORIG_TIER__EUHCount {}
{SEGMENT ORIGINAL SPEAKER EUH PER MIN} SEGMENT_ORIG_TIER__EUHPerMin {format "%.2f" Data}

{SEGMENT ORIGINAL NUMBER OF ELONGATED SOUNDS} SEGMENT_ORIG_TIER__ElongatedSoundsCount {}
{SEGMENT ORIGINAL ELONGATED SOUNDS PER MIN} SEGMENT_ORIG_TIER__ElongatedSoundsPerMin {format "%.2f" Data}

{SEGMENT ORIGINAL SPEAKER NUMBER OF SILENT_PAUSES} SEGMENT_ORIG_TIER__SILENT_PAUSESCount {}
{SEGMENT ORIGINAL SPEAKER NUMBER OF SILENT_PAUSES PER MIN} SEGMENT_ORIG_TIER__SILENT_PAUSESPerMin {format "%.2f" Data}
{SEGMENT ORIGINAL SPEAKER TOTAL DURATION OF SILENT_PAUSES} SEGMENT_ORIG_TIER__TotalDurationOfSILENT_PAUSES_sec {format "%.2f" Data}
{SEGMENT ORIGINAL SPEAKER AVERAGE DURATION OF SILENT_PAUSES PER MIN} SEGMENT_ORIG_TIER__AverageDurationOfSILENT_PAUSESPerMin_sec {format "%.2f" Data}

{SEGMENT ORIGINAL SPEAKER NUMBER OF FALSE_START} SEGMENT_ORIG_TIER__FALSE_STARTCount {}
{SEGMENT ORIGINAL SPEAKER NUMBER OF FALSE_START PER MIN} SEGMENT_ORIG_TIER__FALSE_STARTPerMin {format "%.2f" Data}

{SEGMENT ORIGINAL SPEECH NUMBERS} SEGMENT_ORIG_TIER__CountOfNumbers {}

{SEGMENT ORIGINAL AVERAGE SENTENCE LENGTH (WORDS)} SEGMENT_ORIG_TIER__AverageSentenceLength_Words {format "%.2f" Data}

{SEGMENT ORIGINAL SPEECH WORDS PER THEORETICAL SEGMENT DURATION} SEGMENT_ORIG_TIER__WordsPerTheoreticalSegmentDuration {format "%.2f" Data}
{SEGMENT ORIGINAL UNIQUE WORDS PER THEORETICAL SEGMENT DURATION} SEGMENT_ORIG_TIER__UniqueWordsCountPerTheoreticalSegmentDuration {format "%.2f" Data}
{SEGMENT ORIGINAL EUH PER THEORETICAL SEGMENT DURATION} SEGMENT_ORIG_TIER__EUHPerTheoreticalSegmentDuration {format "%.2f" Data}

{SEGMENT ORIGINAL NUMBER OF SILENT_PAUSES PER THEORETICAL SEGMENT DURATION} SEGMENT_ORIG_TIER__SILENT_PAUSESPerTheoreticalSegmentDuration {format "%.2f" Data}

{SEGMENT ORIGINAL AVERAGE DURATION OF SILENT_PAUSES PER THEORETICAL SEGMENT DURATION} SEGMENT_ORIG_TIER__AverageDurationOfSILENT_PAUSESPerTheoreticalSegmentDuration_sec {format "%.2f" Data}

{SEGMENT ORIGINAL NUMBER OF FALSE_START PER THEORETICAL SEGMENT DURATION} SEGMENT_ORIG_TIER__FALSE_STARTPerTheoreticalSegmentDuration {format "%.2f" Data}

{SEGMENT ORIGINAL ELONGATED SOUNDS PER THEORETICAL SEGMENT DURATION} SEGMENT_ORIG_TIER__ElongatedSoundsPerTheoreticalSegmentDuration {format "%.2f" Data}

{SEGMENT AOT DURATION (SEC) RangeValue} SEGMENT_ANNOTATED_ORIG_TIER__Duration_sec_RangeValue {}

{SEGMENT INTERPRETER SPEECH WORDS COUNT RangeValue} SEGMENT_INTERP_TIER__FullText_WordsCount_RangeValue {}

{SEGMENT INTERPRETER SPEECH WORDS PER MIN RangeValue } SEGMENT_INTERP_TIER__WordsPerMin_RangeValue {}

{SEGMENT INTERPRETER AVERAGE WORD LENGTH (CHAR) RangeValue } SEGMENT_INTERP_TIER__AverageWordLength_char_RangeValue {}

{SEGMENT INTERPRETER UNIQUE WORDS COUNT RangeValue} SEGMENT_INTERP_TIER__UniqueWordsCount_RangeValue {}

{SEGMENT INTERPRETER UNIQUE WORDS PER MIN RangeValue} SEGMENT_INTERP_TIER__UniqueWordsCountPerMin_RangeValue {}

{SEGMENT INTERPRETER AVERAGE SENTENCE LENGTH (WORDS) RangeValue} SEGMENT_INTERP_TIER__AverageSentenceLength_Words_RangeValue {}

{SEGMENT INTERPRETER NUMBER OF EUH RangeValue} SEGMENT_INTERP_TIER__EUHCount_RangeValue {}

{SEGMENT INTERPRETER EUH PER MIN RangeValue} SEGMENT_INTERP_TIER__EUHPerMin_RangeValue {}

{SEGMENT INTERPRETER NUMBER OF ELONGATED SOUNDS RangeValue} SEGMENT_INTERP_TIER__ElongatedSoundsCount_RangeValue {}

{SEGMENT INTERPRETER ELONGATED SOUNDS PER MIN RangeValue} SEGMENT_INTERP_TIER__ElongatedSoundsPerMin_RangeValue {}

{SEGMENT INTERPRETER NUMBER OF SILENT_PAUSES RangeValue} SEGMENT_INTERP_TIER__SILENT_PAUSESCount_RangeValue {}

{SEGMENT INTERPRETER NUMBER OF SILENT_PAUSES PER MIN RangeValue} SEGMENT_INTERP_TIER__SILENT_PAUSESPerMin_RangeValue {}

{SEGMENT INTERPRETER TOTAL DURATION OF SILENT_PAUSES RangeValue} SEGMENT_INTERP_TIER__TotalDurationOfSILENT_PAUSES_sec_RangeValue {}

{SEGMENT INTERPRETER AVERAGE DURATION OF SILENT_PAUSES PER MIN RangeValue} SEGMENT_INTERP_TIER__AverageDurationOfSILENT_PAUSESPerMin_sec_RangeValue {}

{SEGMENT INTERPRETER NUMBER OF FALSE_START RangeValue} SEGMENT_INTERP_TIER__FALSE_STARTCount_RangeValue {}

{SEGMENT INTERPRETER NUMBER OF FALSE_START PER MIN RangeValue} SEGMENT_INTERP_TIER__FALSE_STARTPerMin_RangeValue {}

{SEGMENT INTERPRETER NUMBERS IN INTERPRETATIONS RangeValue} SEGMENT_INTERP_TIER__CountOfNumbers_RangeValue {}

{SEGMENT INTERPRETER SPEECH WORDS PER THEORETICAL SEGMENT DURATION RangeValue} SEGMENT_INTERP_TIER__WordsPerTheoreticalSegmentDuration_RangeValue {}

{SEGMENT INTERPRETER UNIQUE WORDS PER THEORETICAL SEGMENT DURATION RangeValue} SEGMENT_INTERP_TIER__UniqueWordsCountPerTheoreticalSegmentDuration_RangeValue {}

{SEGMENT INTERPRETER EUH PER THEORETICAL SEGMENT DURATION RangeValue} SEGMENT_INTERP_TIER__EUHPerTheoreticalSegmentDuration_RangeValue {}

{SEGMENT INTERPRETER NUMBER OF SILENT_PAUSES PER THEORETICAL SEGMENT DURATION RangeValue} SEGMENT_INTERP_TIER__SILENT_PAUSESPerTheoreticalSegmentDuration_RangeValue {}

{SEGMENT INTERPRETER AVERAGE DURATION OF SILENT_PAUSES PER THEORETICAL SEGMENT DURATION RangeValue} SEGMENT_INTERP_TIER__AverageDurationOfSILENT_PAUSESPerTheoreticalSegmentDuration_sec_RangeValue {}

{SEGMENT INTERPRETER NUMBER OF FALSE_START PER THEORETICAL SEGMENT DURATION RangeValue} SEGMENT_INTERP_TIER__FALSE_STARTPerTheoreticalSegmentDuration_RangeValue {}

{SEGMENT INTERPRETER ELONGATED SOUNDS PER THEORETICAL SEGMENT DURATION RangeValue} SEGMENT_INTERP_TIER__ElongatedSoundsPerTheoreticalSegmentDuration_RangeValue {}

{SEGMENT ORIGINAL SPEECH WORDS COUNT RangeValue} SEGMENT_ORIG_TIER__FullText_WordsCount_RangeValue {}

{SEGMENT ORIGINAL SPEECH WORDS PER MIN RangeValue} SEGMENT_ORIG_TIER__WordsPerMin_RangeValue {}

{SEGMENT ORIGINAL AVERAGE WORD LENGTH (CHAR) RangeValue} SEGMENT_ORIG_TIER__AverageWordLength_char_RangeValue {}

{SEGMENT ORIGINAL UNIQUE WORDS COUNT RangeValue} SEGMENT_ORIG_TIER__UniqueWordsCount_RangeValue {}

{SEGMENT ORIGINAL UNIQUE WORDS PER MIN RangeValue} SEGMENT_ORIG_TIER__UniqueWordsCountPerMin_RangeValue {}

{SEGMENT ORIGINAL SPEAKER NUMBER OF EUH RangeValue} SEGMENT_ORIG_TIER__EUHCount_RangeValue {}

{SEGMENT ORIGINAL SPEAKER EUH PER MIN RangeValue} SEGMENT_ORIG_TIER__EUHPerMin_RangeValue {}

{SEGMENT ORIGINAL NUMBER OF ELONGATED SOUNDS RangeValue} SEGMENT_ORIG_TIER__ElongatedSoundsCount_RangeValue {}

{SEGMENT ORIGINAL ELONGATED SOUNDS PER MIN RangeValue} SEGMENT_ORIG_TIER__ElongatedSoundsPerMin_RangeValue {}

{SEGMENT ORIGINAL SPEAKER NUMBER OF SILENT_PAUSES RangeValue} SEGMENT_ORIG_TIER__SILENT_PAUSESCount_RangeValue {}

{SEGMENT ORIGINAL SPEAKER NUMBER OF SILENT_PAUSES PER MIN RangeValue} SEGMENT_ORIG_TIER__SILENT_PAUSESPerMin_RangeValue {}

{SEGMENT ORIGINAL SPEAKER TOTAL DURATION OF SILENT_PAUSES RangeValue} SEGMENT_ORIG_TIER__TotalDurationOfSILENT_PAUSES_sec_RangeValue {}

{SEGMENT ORIGINAL SPEAKER AVERAGE DURATION OF SILENT_PAUSES PER MIN RangeValue} SEGMENT_ORIG_TIER__AverageDurationOfSILENT_PAUSESPerMin_sec_RangeValue {}

{SEGMENT ORIGINAL SPEAKER NUMBER OF FALSE_START RangeValue} SEGMENT_ORIG_TIER__FALSE_STARTCount_RangeValue {}

{SEGMENT ORIGINAL SPEAKER NUMBER OF FALSE_START PER MIN RangeValue} SEGMENT_ORIG_TIER__FALSE_STARTPerMin_RangeValue {}

{SEGMENT ORIGINAL SPEECH NUMBERS RangeValue} SEGMENT_ORIG_TIER__CountOfNumbers_RangeValue {}

{SEGMENT ORIGINAL AVERAGE SENTENCE LENGTH (WORDS) RangeValue} SEGMENT_ORIG_TIER__AverageSentenceLength_Words_RangeValue {}

{SEGMENT ORIGINAL SPEECH WORDS PER THEORETICAL SEGMENT DURATION RangeValue} SEGMENT_ORIG_TIER__WordsPerTheoreticalSegmentDuration_RangeValue {}

{SEGMENT ORIGINAL UNIQUE WORDS PER THEORETICAL SEGMENT DURATION RangeValue} SEGMENT_ORIG_TIER__UniqueWordsCountPerTheoreticalSegmentDuration_RangeValue {}

{SEGMENT ORIGINAL EUH PER THEORETICAL SEGMENT DURATION RangeValue} SEGMENT_ORIG_TIER__EUHPerTheoreticalSegmentDuration_RangeValue {}

{SEGMENT ORIGINAL NUMBER OF SILENT_PAUSES PER THEORETICAL SEGMENT DURATION RangeValue} SEGMENT_ORIG_TIER__SILENT_PAUSESPerTheoreticalSegmentDuration_RangeValue {}

{SEGMENT ORIGINAL AVERAGE DURATION OF SILENT_PAUSES PER THEORETICAL SEGMENT DURATION RangeValue} SEGMENT_ORIG_TIER__AverageDurationOfSILENT_PAUSESPerTheoreticalSegmentDuration_sec_RangeValue {}

{SEGMENT ORIGINAL NUMBER OF FALSE_START PER THEORETICAL SEGMENT DURATION RangeValue} SEGMENT_ORIG_TIER__FALSE_STARTPerTheoreticalSegmentDuration_RangeValue {}

{SEGMENT ORIGINAL ELONGATED SOUNDS PER THEORETICAL SEGMENT DURATION RangeValue} SEGMENT_ORIG_TIER__ElongatedSoundsPerTheoreticalSegmentDuration_RangeValue {}

{SEGMENT MEAN S2S DELAY (HUNDRETH SEC) RangeValue} SEGMENT_MEAN_S2S_DELAY_hsec_RangeValue {}

{SEGMENT STDEV S2S DELAY (HUNDRETH SEC) RangeValue} SEGMENT_STDEV_S2S_DELAY_hsec_RangeValue {}

}

Appendix 2

Description of the Audio Processing

1. The objective

The objective of the processing of the audio files is to capture the Interpreter Speaker Gender as well as the Interpreer SpeakerID to detect same speakers (note that the Interpreters are considered as anonymous within the EXB files)

2. The input data

The input data are the audio files *.wav, stereo, 24bits, 44100 Hz

3. The pre-processing

The audio files are pre-processed to cope with the requirements of the different tools.
In particular,

- the interpreter channel (the right channel) is extracted
- the format is reduced to 16 Khz, 16 bits

This pre-processing is performed via a batch script FFMPEG_SelectRIGHTChannel_And_ReduceRateAndEncoding.bat using the tool ffmpeg (see hereafter)

4. The processing

The processing is done under the Tcl script “EXBProcess”, selecting the PROCESS_PHASE “AUDIO”.

To support the identification of the SpeakerID, the script has to process a concatenation of audio files (see details hereafter). To limit the volume, the file are truncated, but also split into 6 groups characterized by the Interpreter Sex (M, F) and the Interpreter Language (fr, en, nl).
Those two characteristics could be defined during the AUDIO phase (Sex or Gender is defined by a tool; Language can be deduced from the file name). However, the machine (tool) deduced gender / sex is less reliable than the one manually defined within the EXB files.
Then, a first run of the Tcl script “EXBProcess” with phase “EXB” provides a result file ResultsEXBForAudio_InterpGenderAndLang.txt given for each EXB file a line as follows:
“Normalized File Name” (see hereafter) Gender Language (2 characters)”
(e.g.
“epicg_20080109_common_frame_of_reference_jacquestoubon_i_en F en”

First of all, a normalized file (tail) name is deduced from the original audio file name, to cope with the requirements of some tools but also to be used as cross-reference between the two phases (EXB, AUDIO) of the Tcl script. This normalized file (tail) name is obtained by replacing spaces, “'”, “,” characters by a unique “_” and by lowering the case.

The original files are copied to the renamed files

Thanks to the tool “sox info “, information is retrieved about the audio file to extract the duration of that file.

Thanks to the LIUM_SpkDiarization java script tool, an initial segmentation / diarization is performed and the (sub-)clusters are extracted and characterized from the segmentation result file.

The longest (sub-)cluster is selected and the audio file is truncated to comply with

- a maximal length of 45 sec (parameter Audio_Truncation_Max_Duration_sec)
- the position and length of the selected (sub-)cluster

A segmentation / diarization is performed and the (unique) cluster is characterized from the segmentation result file. The Interpreter Gender (variable "Interpreter Sex Machine") is extracted and appended in a result list.

The lists for concatenation are progressively completed.
A first list contains the (renamed) relative names of the truncated audio files
A second list contains one (initial) segmentation per concatenated file. The start and end segment (counter) are cumulated (that list represents a virtual ideal segmentation of the concatenation audio file)
The concatanation support lists are split into 6 groups (M|F; en|fr|nl).

When all the individual audio files are processed, the script will process the concatenation. For the 6 groups.

The tool "ffmpeg -f concat ....." provides an audio concatenation file
A segmentation / diarization is performed on that audio concatenation file and the resulting segment file is analysed to extract and characterize the (sub-)clusters.

Note that the segmentation limits resulting from the diarization do not exactly cope with the ideal segmentation of the concatenation file.
For every individual file, the common segment lengths between that file and the different sub-clusters are computed. The sub-cluster with the longest common segment length (in fact, the sub-cluster mostly representing the concatenated individual file) is selected and its SpeakerID is extracted.
That SpeakerID is prefixed by the group characteristics (M|F_(fr|en|nl)_) and stored in a result list.

At the end, the audio results lists are saved to be used by the (next) phase "EXB" of the Tcl script

The format of every line is as follows:

- ListOfFileMark_Gender

"Normalized File Name" Gender

e.g.: epicg_20080109_common_frame_of_reference_jacquestoubon_i_en F

- ListOfFileMark_SpeakerID

  "Normalized File Name" SpeakerID (common segment computation results)

  e.g.: epicg_20080109_european_judicial_network_jacquesbarrot_i_en M_en_S107 FileSeg_Length=4498 {{4498 S107}}

The log file contains the processing info as well as the possible WARNINGs.

5. The used tools.

- FFMPEG

The binary tool “ffmpeg.exe” for Windows is extracted from “ffmpeg-20180627-3f95337-win64-static.zip” get at “https://www.ffmpeg.org/download.html” . It is a free software distributed under the GNU General Public License. A documentation is also available. It is a powerful tool to process multimedia stream.

5.2 SOX

The binary tool “sox.exe” for Windows is extracted/ installed from “sox-14.4.2-win32.exe' get at “http://sourceforge.net/projects/sox/files/sox/”. It is a free software distributed under the GNU General Public License. A documentation is also available. It is a powerful tool to process sound data.

5.3 LIUM diarization tool

The main tool supporting the guessing of the speaker gender and speaker (anonymous) identification is the java script based LIUM diarization tool available at “http://www-lium.univ-lemans.fr/diarization/doku.php/welcome”.
It is a complex tool relying, in particular, on various Gaussian based analysis of the audio streams.

It can be freely used at the following condition:
“

If you are using this toolkit in your research please cite one of these papers.

Speaker Diarization

- M. Rouvier, G. Dupuy, P. Gay, E. Khoury, T. Merlin, S. Meignier, “An Open-source State-of-the-art Toolbox for Broadcast News Diarization,“ Interspeech, Lyon (France), 25-29 Aug. 2013

toolkit-interspeech2013.pdf
“

6. On the results.

On Gender recognition

The tool provided a result different from the human defined gender / sex only for 8 files amongst 180 files.

On same Speakers recognition

The tool identified groups of same speakers. The results must be confirmed by human listening.