

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

Collard Camille^{a*} and Bart Defrancq^b

^aDepartment of Translation, Interpreting and Communication, EQTIS, Ghent University, Ghent, Belgium, Camille.Collard@ugent.be, +32495678851, 0000-0002-1938-3828.

^bDepartment of Translation, Interpreting and Communication, EQTIS, Ghent University, Ghent, Belgium, Bart.Defrancq@ugent.be, +32497996274, 0000-0003-0296,0438.

*Corresponding author: Camille Collard, Groot-Brittanniëlaan, 45, 9000 Ghent, Belgium

Camille Collard is a Ph.D student at the Ghent University's EQTIS research group. She graduated as a conference interpreter in 2014 and is conducting corpus-based research on cognitive sex differences in simultaneous interpreting, focusing on Ear-Voice-Span and disfluencies.

Bart Defrancq is an Associate Professor of interpreting and legal translation at Ghent University, and member of the EQTIS research group. He has been and still is involved in several corpus compilation projects, including the compilation of interpreting corpora for simultaneous and dialogue interpreting. He has published widely on contrastive linguistics, translation and interpreting.

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

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

This work was supported by the University of Ghent's *Bijzonder Onderzoeksfonds* (Special Research Fund).

1. Introduction

As part of a broader research project on cognitive sex¹ 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

¹For 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.

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 (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 $\Phi=0.4$ 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 time lag (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 time lag 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 result, 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 automatization 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, Bäckman, & 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

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. Corpus-based studies are sometimes criticized because they consist of samples taken out of the context in which they occurred (Diriker, 2004; Dufrou, 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.

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 at². 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 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 used on the corpus to identify the interpreters' sex. After several necessary

² 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.

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 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 difficulty 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).

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.

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 (R^2), which is a goodness-of-

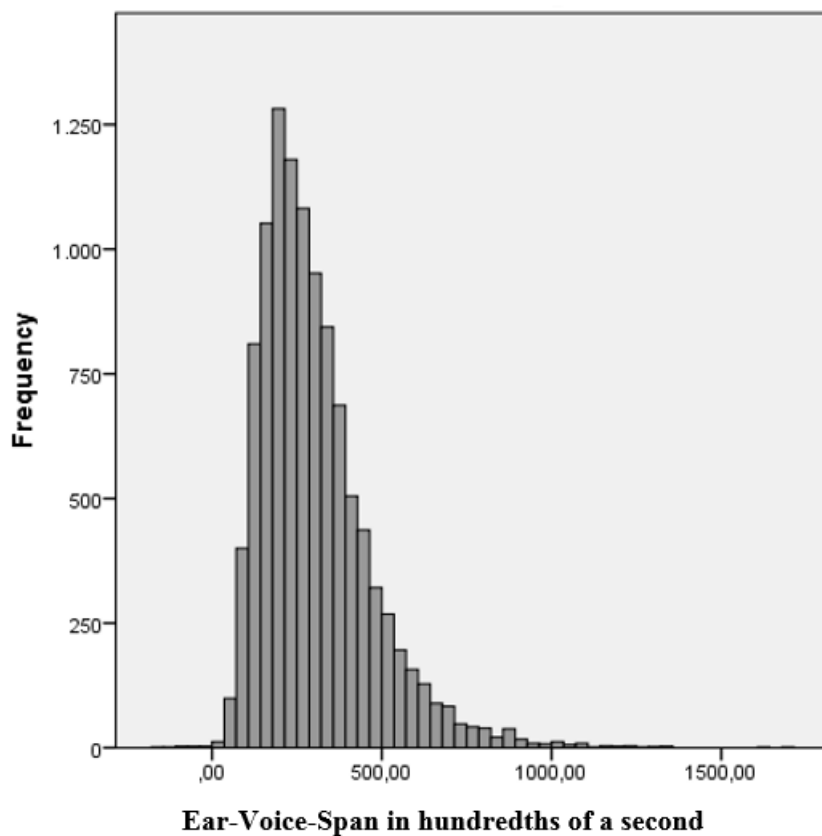
fit measure.

6. Results

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 lengthenings	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

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.

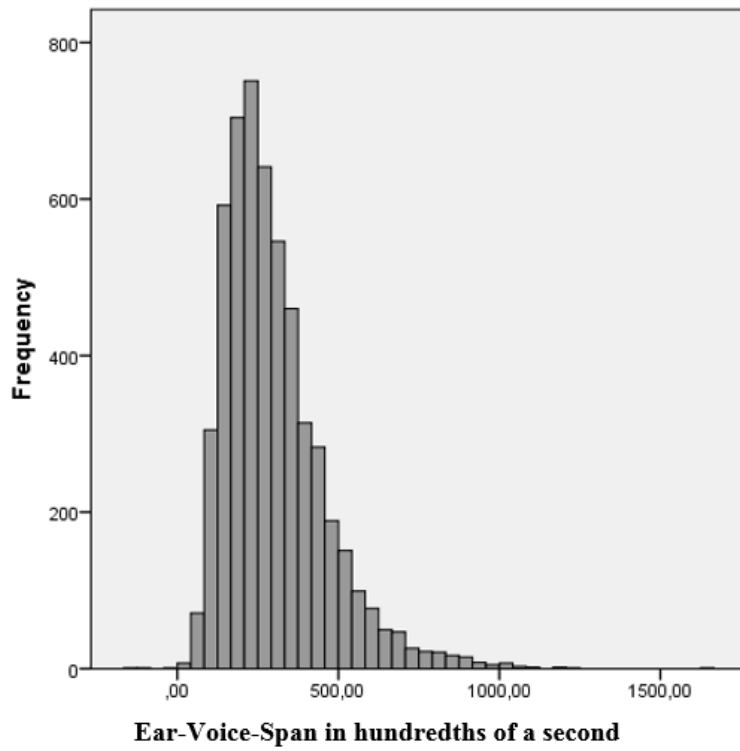
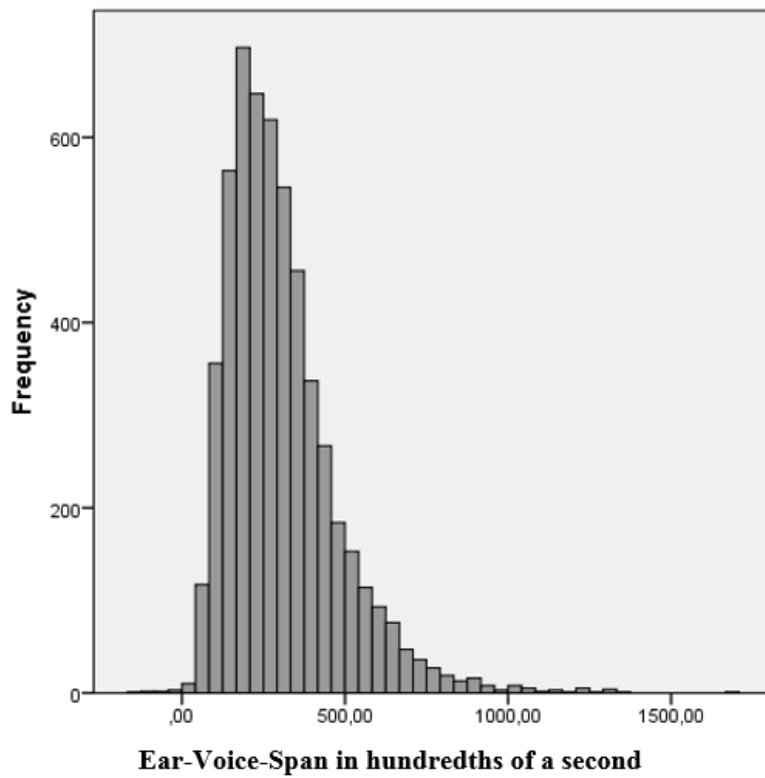


Figure 3. Histogram EVS for male interpreters.



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.

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	<i>p</i> -value	Standardized β	R ²	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 -0.10, 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).

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.

References

- 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.
- 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
- Bachy, S., Dister, A., Francard, M., Geron, G., Giroul, V., Hambye P., Simon, A. C., & Wilmet, R. (2007). 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].
- Baer, B. J., & Massardier-Kenney, F. (2016). Gender and sexuality. In C. V. Angelelli & B. J. Baer (Eds.), *Researching Translation and Interpreting* (pp. 83-96). London: Routledge.
- Baes, E. (2012). *Analyse des différences entre interprètes masculins et féminins au Parlement européen en termes de décalage*. [Analysis of differences in *décalage* between male and female interpreters at the European Parliament.] (Unpublished master's thesis). Hogeschool Gent, Belgium.
- Barik, H. C. (1973). Simultaneous Interpretation: Temporal and Quantitative Data. *Language and Speech*, 16(3), 237-270.
- Bendazzoli, C., & Sandrelli, A. (2005, May). An Approach to Corpus-Based Interpreting Studies: Developing EPIC (European Parliament Interpreting Corpus). In H. Gerzymisch-Arbo, & S. Nauert (Eds.), *Proceedings of the Marie Curie Euroconferences MuTra: Challenges of Multidimensional Translation*. Saarbrücken.
- 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* (pp. 21-42). Singapore: Springer.
- Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glott 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.

- 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.
- 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* (pp. 305-315). Amsterdam: John Benjamins.
- 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.
- 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.
- 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.
- Defrancq, B. (2013, September). *Women and men interpreting*. Paper presented at the first Talking to the World Conference, Newcastle.
- 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* (pp. 25-56). London: Sage Publications.
- 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, PA: 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.

- Duflou, V. (2016). *Be(com)ing a conference interpreter: an ethnography of EU interpreters as a professional community*. Amsterdam/Philadelphia, PA : John Benjamins.
- Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, 40(5), 532-538.
- 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.
- 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* (pp. 52-66). London/New York: Routledge.
- Gile, D. (1995). *Basic Concepts and Models for Interpreter and Translator Training*. Amsterdam/Philadelphia: John Benjamins.
- 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. (2018, January 6) *The effort models and gravitational model: Clarifications and update* [Powerpoint]. 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. (1972). Segmentation of input in simultaneous interpretation. *Journal of Psycholinguistic Research*, 1(2), 127-140.
- Halpern, D. F. (2012). *Sex Differences in Cognitive Abilities: 4th Edition*. 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), 214-8.
- 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. (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, 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* (pp. 51-63). Basel, Switzerland: Karger.
- 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* (pp. 295-304). Amsterdam: John Benjamins.
- 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.
- 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* (pp. 125-144). Berlin: Frank & Timme.
- 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.
- 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.
- 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]* (pp. 147- 168). Granada: Comares.
- Lass, N. J., Puffenberger M. D. (1971). A Comparative study of rate evaluations of experienced and inexperienced listeners. *Quarterly Journal of speech*, 57(1),89-93.
- Lee, T.-H. (2002). Ear Voice Span in English into Korean Simultaneous Interpretation. *Meta*, 47(4), 596–606.
- Lim, H-O. (2005). Meeting Students' Expectations. *Forum 1*(3), 175-204.
- 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* (pp. 159–177). Amsterdam: John Benjamins.
- Loonstra, A., Tarlow, A., & Sellers A. (2001). COWAT metanorms across age, education and gender. *Applied Neuropsychology*, 8(3), 161-166.
- 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.
- 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.
- Mason, M. (2008). *Courtroom interpreting*. Maryland: University Press of America.
- 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.
- Oléron, P., & Nanpon, H. (1965). Research into simultaneous translation. In F. Pöchhacker, & M. Shlesinger (Eds.). *The interpreting studies reader* (pp. 69-76). London: Routledge.

- 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.
- Pöchhacker, F. (2015). Models. In F. Pöchhacker (Ed.). *The Routledge Encyclopedia of Interpreting Studies* (pp. 263–268). New York, NY: Routledge.
- 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>.
- 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*. Interspeech, Lyon.
- Ryan, R. (2015, October 29). *Why so few men? : Gender imbalance in conference interpreting*. Retrieved from <http://aiic.net/p/7347>.
- Schmidt, T., & Wörner, K. (2009). EXMARaLDA: creating, analyzing and sharing spoken language corpora for pragmatic research. *Pragmatics* 19(4), 565-582.
- Setton, R. (1999). *Simultaneous Interpretation: A Cognitive and Pragmatic Analysis*. Amsterdam/Philadelphia: John Benjamins.
- Setton, R., & Dawrant, A. (2016). *Conference Interpreting: A complete course*. Amsterdam/Philadelphia: John Benjamins.
- Shlesinger, M. (1998). Corpus-Based Interpreting Studies as an Offshoot of Corpus-Based Translation Studies. *Meta*, 43(4), 486-493.
- 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.
- 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.
- Timarová, S. (2015). Working Memory. In F. Pöchhacker (Ed.). *The Routledge Encyclopedia of Interpreting Studies* (pp. 443–446). New York: Routledge.
- 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* (pp. 121–146). Amsterdam: John Benjamins.
- 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.