# Evolution of EVS Management in Interpreting Students

Dore, Sara

Sara DORE

# Evolution of EVS Management in Interpreting Students

Mémoire présenté à la Faculté de traduction et d'interprétation pour l'obtention du MA en Interprétation de conférence

Directeur de mémoire : Dr. Kilian G. Seeber

Jurée : Dr. Lucia Ruiz Rosendo

Septembre 2019

Sara DORE

Faculté de Traduction et d'Interprétation

Université de Genève

40, boulevard du Pont-d'Arve,

CH-1211 Genève 4, Suisse

**Abstract**: Ear-voice span (EVS) has often been considered as an indicator of cognitive load in simultaneous interpretation. However, many studies suggest that EVS might be the result of several variables.

The aim of the present study was to observe how EVS management changes as the interpreter acquires experience. The author therefore decided to observe the evolution of EVS over the training of a cohort of interpreting students.

Two Italian students were asked to interpret the same English speech twice, three months apart. The author could therefore observe how EVS changed for each student at two different points of their training. EVS was measured at the beginning of clauses to obtain a mean EVS for each performance. Different categories of word EVS were also measured: keyword EVS (words that the students received in advance), challenging unit EVS, and number EVS. The categories were identified to see whether the students managed EVS differently depending on the nature of the input segment. Omissions in the interpretation were considered as a sign of an unsuccessful EVS.

The results of the comparison between the first and the second interpretation show that EVS did change across time, but less noticeably so across categories.

**Keywords**: EVS, ear-voice span, interpreting experience, training, cognitive load

# TABLE OF CONTENTS

# 1. INTRODUCTION

## 1.1 Overview

Simultaneous interpreting (SI) is an interesting subject for research because of its composite nature. It is a complex cognitive process (Lambert, 1992; Anderson, 1994; De Groot, 1997; Moser-Mercer, 1997; Christoffels & De Groot, 2004) comprising of different tasks, and understanding it is thought to give an insight in human language processing, language production and beyond (Diriker, 2015b). Research in SI is therefore not only the prerogative of interpreters looking to better understand their discipline to improve performance in the profession and to devise better training programmes – this domain is also explored by cognitive psychologists and psycholinguists (Oléron & Nanpon, 1965; Treisman, 1965; Barik, 1973; Christoffels & De Groot, 2004; Köpke & Signorelli, 2012 being just a few examples).

The main tasks which make up the SI process are language comprehension, whereby the interpreter *decodes* the message he/she hears in the source language, and language production, whereby the interpreter *encodes* the message in the target language. Moreover, between decoding and encoding, the interpreter has to store the information, and while encoding he/she monitors the output. These tasks make up a complex process, meaning that many variables can come into play during SI. It is then rather challenging to peel off the layers to understand how SI works, but it appears that ear-voice span can help.

Ear-Voice Span (EVS) is "the delay between the speaker's delivery and the interpreter's output in simultaneous interpreting", in other words "the interval between the interpreter's hearing of what the speaker says and his or her vocal articulation of the interpretation" (Timarová, 2015a:418-419). EVS is one of the rare quantifiable SI variables (Lee, 2006:209) and as such, it is featured in SI studies either as the main topic, or as evidence for other SI factors. EVS is generally held to be the consequence of other elements in the interpretation, usually cognitive load or language structure, but interpretation handbooks (Gillies, 2013; Setton & Dawrant, 2016) encourage interpreters to adopt different lengths of EVS according to the characteristics of the input. EVS management could therefore also be seen as an interpreting strategy in itself. By strategy, we mean a "potentially conscious plan" (Færch & Kasper, 1984:47) aimed at overcoming what can be seen as an obstacle to the interpretation.

The present paper features a summary of the existing literature on the subject of EVS and on related matters, to introduce a discussion on the role of EVS. The author's own hypothesis will be put to a test through a small-scale pilot experiment devised to explore possible changes in students' EVS management between the beginning and the end of their first term of SI training.

### 1.2 Research Question/Hypothesis

Over the course of their training, trainees develop their own way of interpreting, and this may include undergoing some changes in EVS management. EVS length could change as students' cognitive mechanisms develop and as they make different choices, particularly once they have acquired sufficient skills through training and strategies. By acquiring experience in the use of SI strategies, EVS could be shortened to reduce cognitive load in certain segments of the speech and lengthened in others. In the experiment, speech segments are therefore categorised according to their main characteristics, to observe whether they are treated differently from each other and across time in terms of EVS length.

The segment categories are *beginnings of clauses*, *keywords*, *challenging units*, and *numbers*. Beginnings of clauses are used to measure a 'general' mean EVS for each interpretation, which is expected to be brought closer to what is considered to be the average EVS in the literature. Keywords, which were provided to the students before the exercise and therefore should be easily retrievable, are expected to have a short EVS at first, and then have a longer EVS when the students are more experienced. Indeed, after months of training, instead of interpreting the word as soon as they hear it, they could be able to retain the term in their memory in order to devote their attention to more demanding items. Challenging-unit and number EVS are instead expected to decrease over time, as students learn to interpret the most demanding elements as quickly as possible so as to avoid overloading their memory.

# 2. OBJECTIVES AND AIMS

## 2.1 Overall Objective

This study aims at observing the progression of interpreting trainees' EVS management. First, the author gives an overview of existing approaches to the study of simultaneous interpreting and EVS, with the objective of providing a clear picture of the context in which this study situates itself. Through a discussion based on the elements presented in the overview, the author then explains the premises on which the experiment was built. Consisting of two SI exercises performed three months apart by the same interpreting students (within-subject design), the experiment was devised to show whether the students' EVS changes over a term of training.

## 2.2 Specific Aims

Interpreting students are the prevailing subjects of interpreting experiments because they are usually more available and more willing to participate than professional interpreters. When they are involved as a *replacement* for professionals, the results might be skewed because students might not yet have acquired all the abilities required of an interpreter – a disclaimer carefully put forward by the authors whose studies involve student subjects.

This study, however, observes the interpretations of students specifically to discover if and how EVS changes over time during training. To be able to observe this, the author recorded two interpreting trainees interpret the same speech twice: once at the beginning and once at the end of the first of two terms of SI training. The same speech was used so that EVS could be measured in exactly the same points for all interpretations in the same conditions because all source speech variables were kept equal. The only variable that changes is interpreter-related, namely the students' experience, with the exception of a potential 'practice effect'. This effect occurs when changes or improvements in the participants' performance happen because of practice or repetition of the activity being analysed (American Psychological Association, 2018). In this case, the effect may be triggered since the students are called to perform the same activity on the same material twice, potentially confounding the results.

For the purposes of this study, interpreting experience is therefore the independent variable which is expected to affect EVS, the dependent variable. The aim of the author is to observe how experience affects EVS and particularly to determine whether it has different effects on speech segments with different specific characteristics. Such differences could suggest that strategies involving the modulation of EVS are applied.

## 3. BACKGROUND AND SIGNIFICANCE

To better understand the unique nature of SI, it is essential to look at how it has been approached by scholars aiming at shedding light on the processes involved in it. Thus, focussing on the cognitive dimension of SI, the present chapter provides a general idea of what we know so far about this field and in particular EVS.

### 3.1 SI AS A PROCESS

In terms of its use, simultaneous interpreting (SI) is the mode of interpretation that, along with consecutive interpreting, constitutes conference interpreting (Diriker, 2015a). In terms of its nature, SI is an extremely complex cognitive process comprising of several tasks, mainly the comprehension of input and the concurrent processing and production of an earlier input segment (Ruiz Rosendo & Galvan, 2019), placed under "extreme temporal requirements" (MacWhinney, 1997:232).

Few are those who believe that SI is not a cognitively challenging activity, such as Setton (2001), who maintains that if the tasks which comprise the SI process all share the same representation, they can be carried out "comfortably". Indeed, proponents of the Interpretive theory (*théorie du sens*) such as Seleskovitch & Lederer (1986) believe that SI can be performed with standard speech functions. The source language is understood as it would be in any situation, its structure and words are replaced by non-verbal concepts which are then encoded in the target language by the same automated process through which any human produces speech. By virtue of this, the only difficulty posed by SI would be the *simultaneity* of speech comprehension and speech production. However, according to the Information Processing theory, "the surface structure of the message never entirely disappears in the cognitive intermediate processes underlying interpretation" (see Donato, 2003:102). Indeed, most scholars agree that SI is a cognitively complex and demanding process especially because of the *simultaneity* of language comprehension and language production *as well as* the *transposition* of the message from a language into another, as explained by Christoffels & De Groot (2004).

In order to understand how much effort these two aspects require, the SI process has been compared to shadowing and/or paraphrasing tasks (Lambert, 1992; Anderson, 1994; Christoffels & De Groot, 2004) because they comprise of some but not all of SI sub-tasks, which can therefore be analysed separately. Indeed, shadowing and paraphrasing are, respectively, the repetition of the original text, word for word, in the same language, and the reformulation of the original message in the same language, but with different words. If simultaneity of sub-tasks is indeed present in both shadowing and paraphrasing, shadowing does not necessarily require any *understanding* of the content of the utterance (Gerver, 1969). Paraphrasing, on the other hand, requires comprehension in order to express the concept in different words, but it still does not encompass the translation aspect that is part of SI.

The means that have been used to compare simultaneous interpretation, shadowing and paraphrasing, as reported by Christoffels and De Groot (2004), are: quality of performance, ear-voice span (EVS), pupil dilation and Positron Emission Tomography. EVS in this case is considered as the response latency to the given task. All studies show a shorter EVS for shadowing than for SI (e.g. Christoffels & De Groot, 2004), suggesting that the latter is more cognitively demanding. However, there is no way of determining which characteristic makes SI more demanding, since SI has two elements which are not present in shadowing: the task of switching language (code switching) and that of reformulating the message (recoding the input). For this reason, it is useful to bring paraphrasing into the picture, because it does include recoding transformations while it does not include code switching, so a comparison with SI can suggest whether code switching or monolingual recoding has a stronger influence on cognitive load. Christoffels & De Groot (2004) tried to answer this question but could not find a satisfactory answer by comparing SI and paraphrasing. They attribute the mitigated success to the unique characteristics of paraphrasing, which cannot be considered as a monolingual version of SI. However, through the comparison of all three tasks – shadowing, paraphrasing and SI –, something that only Anderson (1994) had done before them, they showed that what makes SI a complex cognitive task is the *combination* of simultaneity and language transformation. Taken each by itself, these two components are less demanding, and the more demanding of the two is the transformation component.

These simultaneous transformations have been studied by several researchers, who have come up with different ways to break down the process of SI into sub-tasks.

Gerver (1970) and Barik (1973) consider the SI process to be composed of 'message reception', 'message decoding', 'target language encoding' and 'delivery into the target language'. On the other hand, Goldman-Eisler (1972) divides SI into three tasks: 'monitoring' of the *input*, 'storing' in the interpreter's memory, and 'encoding' of the message. Davidson (1992) instead maintains the tasks to be 'input', which includes decoding, 'output', which includes encoding, and 'monitoring' of the *output* (not to be confused with Goldman-Eisler's monitoring of the input). By contrast, Christoffels & De Groot (2004) consider the tasks of 'comprehension', 'storage', 'transformation', and 'production' of the segment.

Whichever way we decide to break down the SI process into different tasks, it is important to understand that, while the 'simultaneous' in 'simultaneous interpretation' does indicate that these different tasks take place at the same time, it is not the same segment of speech that concurrently undergoes all of the different steps. Christoffels & De Groot put it quite simply: "The interpreter has to comprehend and store input segments in the source language, transform an earlier segment from source to target language, [and] produce an even earlier segment in the target language" (2004:227).

Researchers also pose the question of *how* the different tasks are carried out together, whether attention is shared equally between the tasks or it is switched between them. No consensus has been found yet on this (Seeber, 2011).

Goldman-Eisler (1972) raises the questions of the order in which the operations are performed and of the relationship between decoding and storing, which should see one of the two being automated and the other receiving attention. Through the study of input segmentation, the author posits that the interpreter stores the segmented input (decoded but not yet recoded according to the target language system) in his/her active verbal memory and retrieves it to recode the message as he/she is encoding it. Thus, the pairs of acts happening simultaneously in the processing of one unit would be monitoring and decoding of the input first, and recoding and encoding of the output second. Since decoding consists of the segmentation of the message, which requires comprehension, the first stage of the process would be the one that requires more attention, while the second one would be more automated. Such is the interpretation that the author gives of what she observed through her study, but she does not provide demonstration for it. The general consensus that in instances where the source text is difficult it is preferable to allocate more attention to the input rather than the output would seem to sit well with this description of the SI process.

Researchers from varying domains, including cognitive psychology and psycholinguistics, bring their contribution to the study of SI because thanks to its complex nature it can help understand human cognitive processes, especially bilingual language processing. At the centre of these studies is the concept of cognitive load, which we have mentioned in passing and which we will expand on in the next segment.

### 3.2.1 Cognitive Load and Working Memory

The concept of cognitive load originates from psychology, it was Gile who introduced it to interpreting studies in 1985, endeavouring to explain interpreter mistakes (Seeber, 2015).

As also reported by Seeber (2013), cognitive load has been considered as the volume of capacity taken up by a cognitive task in a limited capacity system by Miller (1956), as the "mental workload" imposed on the performer by a task by Paas and Merriënboer (1993:739), and as the perceived effort, "the pressure people experience in completing a task" by Yin et al. (2008:2041). By this sequence of definitions, we can notice that the focus has gradually shifted from the task to the performer and his/her perceptions.

Haapalainen, Kim, Forlizzi and Dey (2010) state that perceived effort is the "indicator of pressure on working memory during task execution" (p. 302) and indeed we cannot explain cognitive load without mentioning working memory. The term working memory (WM) was adopted in 1974 by Baddeley and Hitch, when proposing a model for human information processing. They talked of a WM system whose core consists of "a limited capacity 'work space' which can be divided between storage and control processing demands" (p. 76). Over the years, their model was further analysed and expanded by many, including Baddeley himself. In his 2003 review of the model's developments, he states that WM is a system which temporarily stores and manipulates information, and which underpins human thought processes. Also considering other existing WM models (e.g. Cowan, 1988; Ericsson & Kintsch, 1995), Timarová (2015) summarises WM as "a cognitive mechanism responsible for short-term storage, maintenance, and processing of information, and for executive control of cognitive processes" (p. 443). WM does not merely store information, but it can also process it, and it manages attention sharing between tasks – a crucial role in a multitasking process such as SI. Its capacity is limited in terms of the number of discrete items it can store and manipulate concurrently (7±2 items – Miller, 1956), and in terms of time (a couple of seconds – Baddeley & Hitch, 1974). Information can be retained for longer through maintenance, but this requires effort and attention which are ‘taken away’ from other tasks which may need them (Timarová, 2015). In SI, concentrating on the recall of an input unit for longer than usual because, for example, it cannot be immediately encoded in the target language, entails that less attention is devoted to the incoming flow of information, which may then be analysed less thoroughly or even missed or misunderstood. In the worst cases, where information is involuntarily left out in the output, it means that the cognitive load was too heavy and working memory has overloaded, leading to a breakdown in the process. It is therefore essential to be able to manage cognitive load within the limitations of one’s working memory.

### 3.2.2 Cognitive Load and EVS

Seeber (2013) provides a comprehensive description of the measures for the assessment of cognitive load identified by Paas, Tuovinen, Tabbers, and van Gerven (2003): analytical, subjective, performance, and psycho-physiological measures. For the purpose of this study, we will briefly delve into analytical measures, to get an idea of what cognitive load means in SI, and performance measures, to understand why the ear-voice span is so often considered to be the representation of cognitive load.

#### 3.2.2.1 Analytical Measures

A staple among analytical measures is Gile (1985)’s Effort Model, developed to understand recurring difficulties unrelated to linguistic proficiency or other forms of knowledge (Gile, 2015). It divides the complex process of SI into four distinct cognitive efforts: listening and analysis, production, memory, and coordination (Seeber, 2013). Difficulties are therefore explained by a demand of concentration that exceeds available resources or by a mismanagement of these resources. The Effort Model is founded on the theory that all tasks that make up SI exploit the *same* pool of resources.

Seeber (2011) developed instead a more comprehensive Cognitive Load Model based on the premises that each task has a *specific* and finite processing capacity which can be allotted to it, and tasks can interfere with one another to different extents depending on their structural proximity. To calculate this interference, the author uses a conflict matrix where perceptual, cognitive and response processing within the language comprehension task intersect with those of the language production task. The *perceptual* processing of the input will have the highest interference with the *perceptual* processing of the output because they engage the same resources, while it will interfere less with *cognitive* processing of the output, and so on. In the Cognitive Load Model (CLM), the memory or storage component, which is one of the efforts taken into account in the Effort Model, is also added to the load of language comprehension. The CLM can thus illustrate the "moment-to-moment memory load experienced by simultaneous interpreters" and show the "total amount of cognitive load generated by the overlapping component tasks of simultaneous interpreting" (Seeber & Kerzel, 2012:229). It can therefore show how cognitive load fluctuates throughout a performance and where peak load is reached. Seeber (2011) provides visual representations of cognitive load for the four main strategies through which to translate a verb-final language such as German (subject-object-verb (SOV) structure) into a verb-initial language such as English (SVO structure). 'Asymmetrical' language combinations like the above are often taken as the object of cognitive analyses because, seeing as the interpreter cannot follow the original sentence structure, the transformation component is considerable and cognitive load is expected to be higher. Depending on whether the strategy of anticipation, chunking, stalling or waiting is applied, the amount of task-overlapping and interference varies and so do local cognitive load and peak loads.

The CLM models support the position held by Information Processing authors, according to whom "linguistic factors, such as structural asymmetries between SL and TL, are believed to play a significant role in determining an imbalance in the interpreter's mental resources and therefore require language-specific strategies" (see Donato, 2003:102-103). They appear instead to challenge the Interpretive theory, according to which language structure has no effect on SI.

#### 3.2.2.2 Performance Measures

Performance measures in psychological experiments analyse the task in terms of the subject's performance speed and performance accuracy, where speed and accuracy are generally believed to be in a trade-off relationship with each other. "When carried out quickly, tasks will suffer in terms of accuracy, and when carried out accurately, they will suffer in terms of speed" (Seeber, 2013:24). Performance measures observe two tasks performed simultaneously and measure to which extent the secondary task affects the primary task. This is known as a dual-task performance or a secondary task methodology (American Psychological Association, 2018). "The variation of reaction performance [in terms of speed and accuracy] represents the variation in cognitive load" (Haapalainen, Kim, Forlizzi & Dey, 2010). In SI, accuracy is measured through quantifying errors and omissions, while the most direct way of measuring speed, or response latency, is by measuring the time lag, EVS – hence the identification of EVS with cognitive load (Díaz-Gálaz, Padilla & Bajo, 2015).

If EVS were but the response latency to a given task, we would expect it to be shorter when interpretation accuracy is lower, and longer when accuracy is higher. However, several studies show that that is not the case. An interpretation can be accurate even when performed with a short EVS, and long EVS is more likely to lead to loss of content (Barik, 1973; Lee, 2002 and Timarová et al., 2014). Moreover, since simultaneous interpretation already is a multitasking process, it is not possible to observe it by coupling it with a secondary task (Seeber, 2013), and requiring of the interpreters to shift their focus on either the language comprehension or the language production task within SI would mean denaturing the process. Therefore, SI cannot be controlled and manipulated by the researcher as much as an experimental paradigm would require to avoid interferences with the variables being observed. Indeed, Seeber (2013) asserts "In the absence of a traditional secondary task paradigm, establishing a causal relationship between performance speed or performance accuracy and cognitive load at any given point during the process becomes problematic", and adds "Pym [2008] convincingly illustrates how varying priorities among participants can be invoked as an equally plausible explanation of the observed phenomena." (p. 24). Other arguments in favour of the SI process, and particularly EVS, being the result of the interpreter's choices as well as cognitive load are adduced in the following paragraph.

## 3.3 EVS

In SI, Ear-Voice Span (EVS), also known as time lag or *décalage*, is the delay between the beginning of an utterance in the source speech and the beginning of the corresponding utterance in the interpreter's output. It is one of the main features of SI, one of the key elements which describe this mode of interpreting. In fact, EVS is usually mentioned in SI definitions: "Broadly speaking, simultaneous interpreting (SI) is the mode of interpreting in which the interpreter renders the speech as it is being delivered by a speaker into another language with a minimal *time lag* of a few seconds" (Diriker, 2015b:382-383).

EVS has always been part and parcel of research on simultaneous interpreting. Paneth first addressed it in 1957 in what is viewed as the first study on SI, and considered EVS "a characteristic of the simultaneity of interpreting, providing an insight into the temporal development of the target speech in relation to the source speech" (Timarová, Dragsted & Hansen, 2011:122). As seen in paragraph 3.2.2.2, EVS has often been regarded as a representation of cognitive load and has been used to better understand cognitive processes, as it is "one of the few observable and quantifiable variables in SI study" (Lee, 2006:209).

If Lee maintains that "the nature of EVS can be said to be the minimum time for processing on the part of the interpreter" (2006:209), others draw a looser link between EVS and cognitive processing: Christoffels and De Groot (2004), and Timarová (2015a) argue that cognitive factors only determine the 'boundaries' of EVS. Maximum lag being limited by memory capacity, and minimum lag by processing speed (Timarová) or by the smallest amount of information needed to decode the content (Christoffels & De Groot). Díaz-Gálaz et al. (2015) also mention "EVS may tend not to exceed a certain duration, reflecting a natural limit in a simultaneous interpreter's processing capacity" (p. 19). Within these 'boundaries', the duration of the lag can depend on the interpreter's preferences or strategies (Goldman-Eisler, 1972; Anderson, 1994; Lee, 2002, 2006, 2011; Donato, 2003; Kim, 2005; Chmiel et al., 2017).

However, many are the variables over which the interpreter has no control, be them external or internal to the interpreter him/herself. After all, SI is a complex cognitive activity composed of different tasks all happening simultaneously (see §3.2.1) and all of which are influenced by a set of variables.

### 3.3.1 EVS VARIABLES

#### 3.3.1.1 EVS, LANGUAGE COMBINATION AND COGNITIVE LOAD

Much interest has been shown in the influence which language combination could have on EVS and on the SI process as a whole. Some authors, namely those adhering to the *Théorie du sens*, or Interpretive Theory (IT), (e.g. Lederer, Seleskovitch; see Seeber & Kerzel, 2012) are clearly against the idea that language structures may influence processing, because the latter is, to them, universal. However, the hypothesis that language structure does influence the interpreting process seems to be more supported by empirical data analysing cognitive load. Combinations involving structurally and syntactically asymmetrical languages are front and centre because of the perception that those are more difficult to interpret than combinations of more "similar" languages. Among Indo-European languages, German is often studied in comparison to others because of its SOV (subject-object-verb) structure, which appears to have a strong influence on the time lag, because interpreters have to ‘wait’ to hear the verb before encoding the message. The verb is in fact a key element to understanding the meaning of the utterance. In Goldman-Eisler’s words, “the crucial piece of information enabling interpreters to start translation is the predicate” (1972:133); or in Healy and Miller’s figurative words, “the main verb of a sentence defines the plot; the subject merely indicates one of the actors” (1970). Oléron and Nanpon (1965) also suggested that the amount of material which the interpreter needs to hear before he or she can start uttering the output depends on the position of certain key elements such as the verb.

Goldman-Eisler came to her conclusion when a comparison between EVS measurements of German-English, French-English and English-French interpretations showed that “in translating from German, a larger chunk has often to be stored before starting translating than in English and particularly French” (1972:133) because the verb comes later in the sentence. Donato reports similar results in her 2003 study comparing English-Italian and German-Italian SI. She noticed that EVS from English showed a tendency for interpreters to keep their lag as short as possible, while the length of EVS from German depended on the position of the verb, confirming the importance of this element. However, Donato also observed how interpreters of both languages did not rely solely on the verb to start interpreting, indicating that the verb is not as crucial as other studies would suggest. Goldman-Eisler (1972)’s results also showed that the interpreter's subjective organisation of the input, in other words segmentation, depends on the lexical and grammatical features of it.

Bevilacqua (2009) explored the effects of SOV structures (also known as verb-final structures) on the strategies of interpreters working from two SOV languages, German and Dutch, into Italian, a verb-initial language, and showed the important role that anticipation has in interpretation between SOV and SVO languages.

When anticipating, the interpreter infers from the other constituents of the sentence what verb the speaker is going to use and therefore does not feel the need to wait until the verb is uttered to start interpreting that sentence. Anticipation therefore shortens the time lag and is a way to avoid the cognitive load that comes with keeping the whole sentence in the working memory, although the effort of anticipation could bring its own share of cognitive load. Anticipation is one of the strategies used to deal with asymmetrical language structures which several authors have studied (Kirchhoff, 1976 and Kalina, 1998, quoted in Donato, 2003; Kohn & Kalina, 1996; Setton, 1999) and which Seeber thoroughly analysed together with waiting, stalling and chunking in terms of the cognitive load they amount to (2011; Seeber & Kerzel, 2012). The choice of which strategy to apply affects the length of EVS and shifts the peak of cognitive load to different stages of the SI process.

Ono, Tohyama and Matsubara (2008) ventured that the difference of verb positions between Japanese and English might be at the source of the longer-than-average EVS they measured in Japanese to English interpretation. Conversely, when Kim (2005) compared interpretations from Korean, an SOV language, to Japanese, another SOV language, and to Chinese, an SVO language, surprisingly there was no significant difference in EVS in the two combinations, despite the different language structures. The author suggested that the reason for those results were the different approaches chosen by the interpreters to deal with the contrasting requirements of the language combinations. Further, Kim not only considers the differences in the syntactical structures of the languages, but also the number of syllables that the languages involved need to express the same concept. Hence, she explains how EVS is shorter in Korean-Japanese SI than in Japanese-Korean SI because Japanese has a higher syllable count, so the Japanese interpreters not only speed up their utterance, but also adopt "the strategy of reducing the time taken for information processing and allotting more time for utterance" (2005:12), and to do so they need to start talking as soon as possible.

As mentioned earlier, syntactically symmetrical language combinations are regarded as easier to interpret because they "demand less syntactic restructuring than do languages which differ considerably in structure" (Donato, 2003:128). Nonetheless, in an analysis of SI between cognate languages through the example of Spanish-Italian interpretation, Fusco (1990) mentions that the similarity between the two languages can interfere with the quality of the output. The source language can influence the syntax, the morphology and the lexicon of the target language by activating similar but incorrect elements. For instance, if a noun is very similar in the two languages but has a different grammatical gender, the interpreter might easily enough avoid the mistake of misgendering the noun itself, but will require increased awareness to make all attributes agree with it, especially if they are separated by other constituents. The attention required to suppress the wrong grammatical gender feeds into the cognitive load and might slow down the response.

Different language combinations therefore seem to influence the processing of information to different extents, whether it is shifting the focus on reformulation of the message or on suppression of misleading input. Interpreters can adopt different strategies depending on the requirements of the languages with which they are working, and these strategies can in turn affect EVS or consist of modulating EVS.

However, many other factors have been found to interact with EVS, whether they influence it, or are influenced by it.

#### 3.3.1.2 EVS AND SOURCE TEXT FEATURES

Barik (1973), De Groot (1997), Lee (2002) and Ruiz Rosendo and Galvan (2019) explored the effects of source text speed on EVS, with different results. On the one hand, Barik and De Groot found that a higher input rate results in longer EVS, probably because a high rate makes the message denser and therefore, according to Mazza (2001), more difficult to process.

Davidson's (1992) study is in line with these results because he notices that high input rate is one of the instances where students lag further behind; although the fact that they are students leaves room to believe that with experience they could develop a strategy to decrease their lag when facing a fast-paced speech. Timarová et al. (2015) believe indeed that more experienced interpreters have shorter EVS because of strategies and automatization (Chmiel et al., 2017). Along the same lines, Lee found that Korean interpreters shortened their EVS when the English input was faster in a deliberate attempt to avoid overloading their Working Memory. Instead, Ruiz Rosendo and Galvan found no significant correlation between source text speed and EVS length in English into Spanish interpretation of a specialised discourse.

Lee (2002) explored other 'speaker variables' such as speech/pause ratio, between-sentence pauses and sentence length, and their effects on EVS. According to his findings, a high speech/pause ratio in the original speech meant that the interpreter would use a shorter EVS than when this ratio decreased. Long sentences were interpreted with a long EVS, while shorter sentences had a shorter EVS. Long between-sentence pauses, instead, lead to a shorter EVS than short pauses did. Lee also claims that EVS length affects in turn 'interpreter variables'. For instance, a short lag leads to a higher speech/pause ratio in the interpretation.

On a more global level, source text type (structured or impromptu) can also influence the time lag, as Oléron and Nanpon (1965), Barik (1973) and Adamowicz (1989) found out. Having access to the text in its written form to perform SI with text appears to lengthen EVS (Lamberger-Felber, 2001). The level of complexity of the speech (Schweda-Nicholson, 1987; Díaz-Galaz, 2011), especially in relation to the interpreter's knowledge of the subject, is also a determining factor of EVS because it can require more or less cognitive effort. The more information is already stored in long-term memory, the less working memory will be loaded because the interpreter can anticipate what the speaker is going to say.

#### 3.3.1.3 EVS AND THE INTERPRETER

For the same reason, advance preparation of the subject to be interpreted usually facilitates the cognitive process and shortens EVS (Díaz-Gálaz et al., 2015).

Collard and Defrancq (2019) conducted a study to establish whether the sex of the interpreter is a predictor of EVS, seeing as women and men do perform differently in single cognitive tasks which are part of SI, but came up with no statistically significant difference between the sexes.

Directionality – whether the interpreter is working into his/her mother tongue, the A language, or a foreign, B language – is also likely to affect cognitive processes and EVS, because suppression of the input and activation of the output in these two cases can work differently (Seeber, 2017).

The interpreter's own cognitive mechanisms, aptitudes and preferences can also have an effect on the time lag, according to Anderson (1994). Several sources (see Collard & Defrancq, 2019), among which Timarová et al. (2011), report that EVS varies notably not only within one performance of the same interpreter, but also among the performances of different interpreters. Such results seem to support Anderson's claim. If, for instance, an interpreter needs to pay particular attention to the listening part of the process because the source language is his/her weakness, he/she will probably have a longer EVS. Since both a short EVS and a long EVS can have their own drawbacks (following too closely the source language structure and having to correct oneself, and losing content, respectively), the interpreter may have a preference for running one risk rather than the other. Indeed, Kim (2005), Lee (2006), Chmiel et al. (2017) and Ruiz-Rosendo and Galvan (2019) consider that EVS can be a strategy in itself.

*3.3.1.3.1 EVS AND INTERPRETING EXPERIENCE*

Another variable which is commonly held to affect time lag is the interpreter's professional experience, although Díaz-Gálaz et al. found no correlation between EVS and experience in their 2015 study. Among those who have found a correlation, there is no consensus as to how experience affects EVS. Davidson (1992) maintains that novice interpreters are "anxious to deal with the speaker's output" (p. 5) and therefore tend to translate lexical phrases before they have enough information for disambiguation, while more experienced interpreters wait to hear more before they decide whether or how to translate the phrase. Moser-Mercer (1997) equally states that expert interpreters usually have longer EVS because they have a more comprehensive view of the message. On the other hand, Timarová et al. (2015) found that more experience meant a shorter time lag, and Ruiz Rosendo and Galvan (2019)'s results suggest that experienced interpreters have a shorter EVS than novice interpreters, especially at high input delivery rates. However, such difference is not statistically significant.

A logical explanation can be found for either scenario. Trainee interpreters might be inclined to wait longer to get a bigger picture, especially if they are new to SI practice and used to training in consecutive interpretation, which usually precedes SI in training programmes. Indeed, in consecutive interpretation, in order to remember the message and be able to deliver it several minutes after the beginning of the speech, they need to perfectly understand the logic of the whole speech and they are able to work with a much bigger chunk of information. Trainees might equally have a longer lag because they are not as comfortable performing all the different tasks involved in SI at once, hence their processing takes longer. On the other hand, experienced interpreters are more likely to have shorter response latencies because of automatization and/or strategies that they have developed through training and working.

However, the fact that they can potentially wait less because they have fewer restraints on working memory does not necessarily mean that their EVS will be shorter – they might prefer otherwise, as shown by Davidson's results. A lighter cognitive load might simply mean that interpreters can make more deliberate choices. As already mentioned, Davidson also points out that novice interpreters might be anxious not to miss information so they will wait very little and translate almost word for word. A more literal translation meaning that less effort is put in the translation task, overall cognitive load might consequently be lower, allowing for a shorter EVS.

Which approach the trainee interpreter adopts probably ultimately depends on their personality and their cognitive aptitudes. No matter the starting point, all interpreters aim at proficiency, developing their own techniques and trying to maintain optimum EVS.

### 3.3.2 OPTIMUM EVS

Nonetheless, determining what EVS is optimum EVS is a complex matter. As argued by Chmiel et al. (2017), interpreting students are encouraged to lengthen their EVS so as to avoid the aforementioned word-for-word interpretation which is likely to make their output sound unnatural or even require a "costly correction" (De Groot, 1997; Lee, 2002). However, Defrancq (2015) showed that short EVS is not necessarily linked to lower SI quality, and according to Barik (1973), Lee (2002) and Timarová et al. (2014) it can even lead to higher accuracy and more completeness. Lamberger-Felber (2001) and Lee (2002), among others, even report long EVS being responsible for loss of content, affecting not only the concurrent item, but also the following one. De Groot (1997) and others advise to keep as short a time lag as possible. However, EVS can change within a single speech, and a short or long lag might be preferable according to the characteristics of different segments of the speech. Gillies (2013) and Setton and Dawrant (2016), for instance, suggest short EVS is necessary to avoid omissions when interpreting non-redundant segments such as names, numbers and enumerations. Numbers are particularly demanding because they are not only characterised by low redundancy (Gile, 1984), but also have high informative content (Alessandrini, 1990), low predictability (Braun & Clarici, 1996), and are difficult to remember (Mazza, 2001). It is therefore necessary to have enough resources available to understand numbers correctly and encode them before working memory overloads. Ono et al. did indeed find out in their 2008 study that "numerals are interpreted quickly regardless of the language pairs" (p. 3387), as did Timarová et al. (2011). All this suggests the existence of a strategy by which interpreters try to incorporate numbers in their outputs as soon as possible to avoid burdening their working memory with such difficult segments.

Ultimately, optimum EVS would appear to fluctuate even within a same speech, according to the requirements of the segments that compose it and to the interpreter's cognitive approaches to said requirements.

### 3.3.3 MEASURING EVS

EVS has been measured in terms of time – seconds, centiseconds or milliseconds – (from Paneth, 1957; Oléron & Nanpon, 1965; Barik, 1973 to Kim, 2005; Díaz-Gálaz, Padilla & Bajo, 2015 and Collard & Defrancq, 2019 among others), lexical/syntactical units (Goldman-Eisler, 1972; Adamowicz, 1989; Davidson, 1992; Donato, 2003) and words (Treisman, 1965; Schweda-Nicholson, 1987).

While all the authors had good reasons for choosing the unit of measurement they chose, the lack of homogeneity hinders comparability in some cases. However, since many studies measure EVS in seconds, a general consensus could be found around average EVS being of two to four seconds in SI between Indo-European languages.

Lee (2002) suggested that a lag longer than four seconds is more likely to lead to a loss in quality. Studies on SI involving oriental languages sometimes yielded different results: Ono, Tohyama and Matsubara (2008) found an average EVS of 2.446 seconds for English into Japanese SI, but in Japanese into English SI the mean was 4.532 seconds, while Kim (2005) found averages between one and two seconds for Japanese into Korean, Korean into Japanese, Chinese into Korean and Korean into Chinese. If Oléron and Nanpon (1965) found that EVS generally varied from two to 10 seconds and Lederer (1978) measured lags from three to six seconds, more recently, in a 2015 study on a corpus of real-life interpretations, Defrancq found that EVS could span from -1 to 10 seconds. As he explains at page 31:

> A negative EVS implies that the interpreter is ahead of the speaker and is anticipating what the speaker is about to say […] Where EVS is 0 seconds, 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 (let alone process) the source language input and must therefore, in practice, have anticipated it.

*Where* to measure EVS is an important choice researchers need to make, because EVS values themselves are affected by it. As we have seen previously, some languages place their verbs later on in the sentence than others. Therefore, if EVS were to be calculated for verbs only, values would be skewed compared to the bulk of the interpretation. In a German to Italian interpretation, EVS would probably turn out to be shorter between verbs because the Italian interpreter is likely to encode them as soon as possible, to accommodate the requirements of his/her language. The same can happen with other sentence constituents which are best suited in different positions according to the language. It is to even out these possible outliers, that Christoffels and De Groot (2004) measured the EVS of task sentences by averaging EVS values for three content words from the beginning, the middle and the end of the sentence. In their study, they worked on the interpretation of discrete sentences and calculated EVS for half of them, while Collard and Defrancq (2019) worked on a speech paradigm, so they measured average EVS for the 10-second segments into which they had divided the input. The average EVS for each segment was calculated through measurements taken every 10 items uttered in the output. Ono et al. (2008) take content words as cues in the input, but they consider that the time lag begins at the *end* of these words and ends at the beginning of the corresponding output.

Defrancq (2015) measured EVS to the nearest second by applying time tags to equivalent lexical items, even when a mistranslation occurred. When the different structures of the two languages (French and Dutch) meant that word order would be different within a syntactic unit, he measured time lag between the two closest items within the units.

Donato (2003) measured EVS at the beginning of each sentence in terms of which syntactic segment is necessary for the interpreter to start encoding the message, an approach previously adopted by Goldman-Eisler (1972). Díaz-Gálaz et al. (2015) as well as Ruiz Rosendo and Galvan (2019) measured EVS at the beginning of each input segment. Kim (2005) does not specify whether she applies time tags at the beginning of each segment of the real-life interpretation, but it would appear so from the ensuing comments on results. She gives the average of all EVS values for each language combination, then she excludes cases where the content could be encoded in under one second "such as when the adverb indicating tense appears at the *beginning of the sentence* or when conjunctions make it unnecessary to divide into units of meaning" (p. 6).

Lamberger-Felber (2001) only measured EVS in instances of considerable omissions, to explore a possible relation between long EVS and loss of content. Different cues can therefore be chosen as a starting point for the time lag to be measured, depending on the variable that is being explored.

Lee (2006:209) proposes a "more accurate variable than mere EVS": "EVS/sentence ratio", not too dissimilar in principle to Goldman-Eisler's "EVS units" (1972) in that they both put EVS in relation to a more immediate context, the sentence. While Goldman-Eisler considers the sentence in its syntactical components, describing EVS as a sum of one or more of those components, Lee takes the sentence as whole, expressing EVS as a fraction of the sentence's length: "*When EVS/sentence is 1, the interpreter begins to produce converted information after listening to the speaker's entire sentence*" (p. 209).

Another less famous variable exists: the tail-to-tail span, which measures the time-lag between the end of the speaker's segment and the end of its interpretation (Lee, 2003). Such a variable could be useful in the study of concision and in determining how much of a segment's EVS depends on the way the previous segment was interpreted. However, EVS is the variable used for the purposes of this study, in the modalities which are described in the following chapters.

## 4. RESEARCH DESIGN AND METHODS

Based on the knowledge of the existing studies presented in the previous chapters, a small-scale pilot experiment was devised to monitor changes in the ear-voice span (EVS) of interpreting trainees at the beginning and at the end of the first term of formal SI training. The aim of the exercise is to detect a pattern to suggest that EVS changes as experience increases. The author believes that, while cognitive load is a relevant factor in SI due to the overlapping of several tasks, its influence decreases as the interpreter gains experience and partly automatizes the process. As the restraints imposed by cognitive load have decreased, the interpreter may have more leeway to manage his/her own discourse and EVS.

To test whether EVS was affected by experience, two interpretations of the same student, performed at two different times of his/her SI training (part I and part II) needed to be compared. The author chose a *within-subject design*, comparing performances from the same group of trainees across time, rather than comparing a less experienced group of trainees to a more experienced group of trainees (Davidson, 1992), to observe how the same person adapts to his/her own cognitive aptitudes. To decrease confounding variables, the source speeches for both part I and part II had to be as similar as possible. As EVS can be affected by many variables in the input (speed, sentence structure, cognitively demanding segments, etc. see §3.3.1), using the same speech twice appeared to be the easiest way to ensure that these elements would not interfere. As Setton pointed out, "empirical investigation is hampered by problems of multivariate comparability across languages, discourses, events and interpreters" (1999:55).

This choice does, however, possibly entail the interference of another variable: familiarity with the task (practice effect). Even if the students practise on dozens of different speeches in the course of a few months, and even if they do not try to actively remember the source speech after having interpreted it the first time, their long-term memory could be triggered once they hear the speech a second time. Remembering anything from the logical sequence of the utterance to the more challenging segments could make them more aware and shorten their reaction time.

Nevertheless, the advantages of using the same speech outweighed the possibility that students may partly remember it. Some interpreting trainees enrolled in Conference Interpreting training programmes with similar curricula were therefore invited to take part in the experiment as volunteers; two agreed to take part. They were asked to interpret a six-minute speech from English, a C language, into Italian, their A language. The first time, they interpreted it while in their second week of SI training; the second time, they were in the 14th and last week of the same academic term. At that point, their SI training was not yet complete, as they still had to go through a period of independent training and a second and final term of training.

The original speech and the interpretations were simultaneously recorded on separate tracks each time. Different segments of the original speech were then identified according to the functions that made them relevant to EVS measurement, and time tags were manually applied on all tracks to obtain those measurements. An analysis of this data and of the transcriptions is the basis for the discussion of the hypothesis.

## 4.1 Population and Study Sample

This experiment observes the progression of EVS values within the training of aspiring interpreters. As such, any interpreting student with any language combination at the beginning of their formal SI training could be considered part of the theoretical population. Indeed, language combination does affect EVS, but the main purpose of the present study is not to compare performances by different students, instead, it looks for noticeable EVS changes over time *within* the same subject's interpretations. With all required considerations for each language combination, a comparison could therefore be made of two performances for each interpreting student to establish whether their own EVS changes over time.

However, the author chose to collect Italian interpretations of an English speech because she is herself trained in English into Italian interpretation, them being her B and A languages respectively. This choice should allow her work to be more effective. The study sample therefore consists of two Italian-A, English-C interpreting trainees who had just started their first term training in SI. When performing in part I, they had both already trained in consecutive interpreting for a term, and they were starting the first of two terms of SI training.

Two subjects make up a rather small sample, which could not be representative of the whole population of interpreting trainees. For this reason, the analysis is of a merely descriptive statistical nature. The results could then suggest that an experiment on a larger scale may teach us something more about EVS.

## 4.2 Sources of Data

### 4.2.1 Materials

The speech was chosen from a pool of pedagogical videos not accessible to students, so that they would not chance upon it. The source file consists of a six-minute video where the speaker looks directly at the camera and clearly enunciates the speech in standard American English. The students could therefore see the speaker, which is the usual condition in which they train, and generally the preferable option for any interpreter. The actual speech only started after the speaker had stated the title, which the students were not required to interpret, and it was uttered at a speed of 123 words per minute. Because of the restricted number of available speeches, the level of difficulty of the chosen one, assessed according to the standards of the Speech Repository of the European Commission (n.d.), was higher than what beginner students are expected to interpret. This could pose some difficulty with regards the interpretation in part I of the experiment.

### 4.2.2 Procedure

To analyse and compare the interpretations of the English speech, they first had to be recorded.

The students were therefore invited to participate in the experiment during their second week training in SI. They received a letter (see appendix 1) detailing what they would be asked to do and the rules of confidentiality. For those who accepted, the date of part I of the experiment was set for the end of their second week of SI practice, meaning that the double task of listening and speaking simultaneously would not be so novel as to overwhelm them, but they were still at a beginner level. The whole experiment was to take place at the interpreting department of the University of Geneva.

The subjects were welcomed in one of the rooms available to interpreting students, equipped with booths and the recording program Sonus - Televic Education. Following a previously-set-up checklist (see appendix 2), the author made sure that the students knew how the equipment worked, and briefly explained them the nature of the exercise and when they should start interpreting. The students were not told what the analysis would focus on, nor did they know that they would interpret the same speech in the second part of the experiment. The students were asked to limit unnecessary noises in the booth (e.g. paper rustling) and to keep interpreting throughout the exercise.

The author then provided each student with a list of keywords to prepare the speech for 10 minutes (see appendix 3). Since the speech was of an intermediate level of difficulty, the author felt it important to give the subjects enough time to familiarise themselves with the topic and the vocabulary.

Both students then took their places in two separate booths and, after a brief sound-check, the exercise started. They interpreted the six-minute speech while the author monitored them from the same room. Televic Education recorded their interpretations and the original speech on separate tracks, creating two files with the original speech as Track 1 and either one of the interpretations respectively as Track 2.

At the end of the exercise, the author thanked the students and asked them to return the list of keywords. It was a necessary measure to lower the chances that they may remember the content of the speech, since they would have to re-interpret it three months later.

The original speech and the interpretations were transcribed to identify the categories of words and the corresponding items in the output text.

The same protocol was followed for part II of the experiment, which took place on the 14th week of training at a similar time of day, in another room available to interpreting students. The subjects did not have any reason to suspect that they were interpreting the same speech as in part I, until they recognised the keywords they were given to prepare for 10 minutes.

## 4.3 Collection of Data

The experiment resulted in four audio-visual files, two for each part of the experiment, and two for each participant. These files were converted into audio-only files for use on Audacity. The original speech was transcribed, and the salient parts for EVS measurement were identified in the text. Time tags for EVS measurements were applied manually at the onset of each of these elements on the speaker track and at the onset of their corresponding elements on the interpreter track of each file. The length of the ensuing segments was automatically calculated by Audacity in seconds. EVS is measured in seconds to facilitate comparison with existing literature.

It was decided to measure EVS at the beginning of input clauses to gauge the subject's general tendency for a longer or shorter time lag, and whether it changes from part I to part II. In order to be chosen as starting points for EVS measurements, the segments had to be beginnings of clauses preceded by a pause (silences over 250 milliseconds, as suggested by Goldman-Eisler, 1972). Segments which could be considered as clauses but were not explicitly separated from the rest of the text by pauses were excluded from the count so that the author's own segmentation would not interfere. Instances where pauses separated elements within the same syntactic unit were not taken into consideration either. The time tag on the interpreter track was applied at the onset of the unit translating the beginning of the input clause even when it started by a different word.

Corresponding words were instead marked to measure EVS for elements expected to require a particular strategy (i.e. numbers and challenging units – see appendix 4), and for words expected to be already activated because they were given in the preparation (keywords). We will refer to these words as 'words of interest'. The 'challenging units' category is made up of technical terms selected by the author which were not given in preparation and which are not used in common parlance.

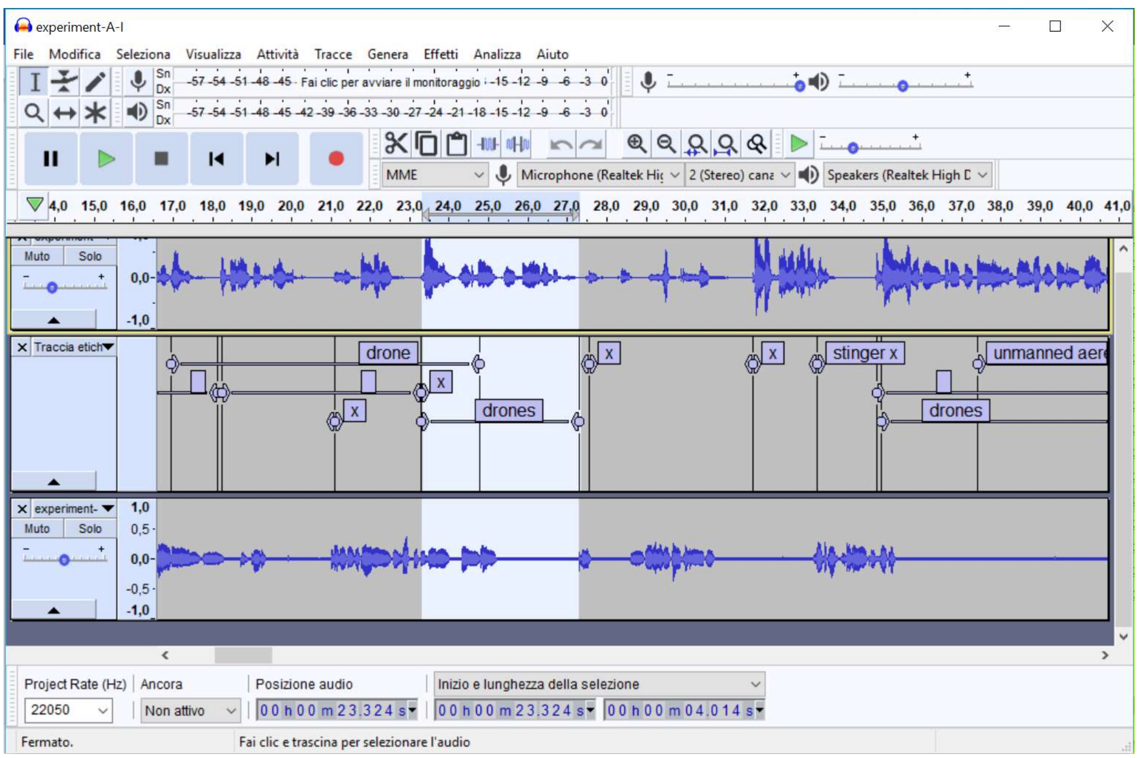
As in Defrancq (2015), sometimes, corresponding elements to the input word were mistranslations, but the EVS value was still considered valid. Pronouns replacing the word of interest in the output were also used for EVS measurements. When the translation of a relevant element was omitted, EVS could not be measured and was therefore considered null. When word order between the two languages was different within a syntactic unit, EVS was always calculated at the onset of the first word of the unit; e.g. "**bird**'s eye-view" translates to "vista a volo d'**uccello**". Although the Italian for "bird" is "uccello", the time tags were positioned before "bird" and before "vista" because they mark the beginning of their lexical unit.

In cases where the word of interest is repeated twice in the input speech and only uttered once in the interpretation, the EVS measurement only takes into account the first for simplicity, because there is no way of knowing whether the processing of the word was triggered by the first utterance, or if the interpreter would have missed the word had it not been repeated. Another way to see this would be considering the repetition as an utterance of its own that is omitted in the interpretation. The first word would therefore have the measured EVS and the second word would have a null EVS, as is the case for all omissions.

## 4.4 Data Analysis Strategies

Each interpretation was analysed on Audacity as a two-track recording. Time tags were applied to salient segments in the source speech first, and then to the corresponding items in the interpretation, if uttered. While time tags were added manually through auditory and visual cues (see fig.1), the measurement of the time lags was automatically provided by the program. Once collected, these values were processed on Excel.

Fig.1: screenshot of Audacity during measurement of EVS




EVS values from part II were compared to data collected in part I according to their category (keywords given for preparation, numbers, challenging units, and beginnings of clauses). An average EVS was calculated for each performance by each student, using the time lags measured at the beginning of clauses. Word EVS values were considered likely to skew the calculation because of differences in word order between source and target languages, hence their processing as a separate element. A comparison was made between the resulting EVS averages for the two interpretations of the same student to spot potential general differences in EVS length. As to word EVS, averages were compared within the same category. The aim was to observe whether there were changes suggesting that students were learning to approach the different categories of words of interest with a specific strategy, possibly resembling those indicated by textbooks.

## 4.5 Ethics

All data gathered during this experiment was dealt with confidentially, participants remain anonymous and are referred to as 'student A' and 'student B'. After receiving a letter detailing the nature of the experiment and what was required of the subjects, they agreed to participate on a voluntary basis. Everything was organised to best fit their availabilities and to make them comfortable in the experimental environment.

## 4.6 Timeframes

The experimental set-up took about four months: one week to select the source speech with the right characteristics and find an agreement with the voluntary subjects, three months for data collection – although the proceedings involving the subjects took less than one hour overall –, and one month for data analysis.

## 5. DATA AND DATA ANALYSIS

The recordings of the interpretations are named in the following manner: student A's performances are A1 for the first part of the experiment and A2 for the second part, and student B's performances are called B1 and B2 for part I and part II respectively.

A total of 151 measurements were taken for each performance. Of these, 82 were beginnings of clauses, 57 were keywords given in the preparation (the preparation sheet consisted of the title of the speech and 19 terms – see appendix 3 –, but since some of them were repeated several times during the speech the count is higher), nine were challenging units (seven different elements, one of which was repeated three times – see appendix 4) and three were numbers (see also appendix 4).

Average EVS was calculated for each of these categories. When EVS could not be calculated because the unit did not appear in the output, the time tag was excluded from the calculation of the average, and the author kept count of the omissions instead. While the present study does not focus on interpretation quality, omissions are taken into account as an indication of quality in terms of completeness, as a factor to map onto EVS patterns to understand whether they are effective. If many omissions are detected at a specific point of the interpretation, it stands to show that the kind of EVS employed there did not deliver.

The average calculated for EVS at the beginning of clauses is here considered to represent the general EVS of the interpretation because it concerns the translation of whole segments, while other EVS measurements concern words whose positions could vary according to the language. The standard deviation (SD) was calculated for beginning-of-clause EVS and keyword EVS to quantify the dispersion of the values and exclude outliers (values over or below the range of two standard deviations) from the count. Outlier results fell strictly above the average EVS and did not exceed three per group. This operation was not applied to the two remaining groups due to the small number of elements of which they comprise.

Word EVS categories are compared to verify whether units with different characteristics are interpreted with different EVS lengths. Number EVS is expected to decrease because numbers present a heavy burden on working memory and it is therefore preferable for interpreters to encode them quickly. Challenging unit EVS is expected to be lower in part II, while keyword EVS is expected to increase. In fact, if EVS is not merely a latency measure, easily retrievable terms such as those prepared beforehand should not necessarily be encoded as soon as they are activated in the interpreter's vocabulary. The interpreter could decide to encode them later in the output because they are a light burden for working memory, instead concentrating on encoding challenging units more quickly.

For a more accurate comparison between A1 and A2 and between B1 and B2, another set of averages is provided, which takes into account only the items that are interpreted in both performances by the same student – we will refer to this set of data as 'intersection' data.

The main features observed for each performance are listed below, establishing the starting point for the comparison among performances. The EVS values to which the following paragraphs refer are listed in appendix 5.

## 5.1 Part one

### 5.1.1 Performance A1

The general average EVS for A1, calculated through 47 valid *beginning-of-clause* time lags, is 5.47 seconds, with a standard deviation SD=2.56s. The average calculated on 45 values excluding the two outliers is 5.17s. Although words that are already activated through preparation are expected to be interpreted faster, the average time lag for *keywords* is 6.48s, excluding the single outlier (SD=2.57), while *challenging units* were interpreted on average after 5.90s. Activated terms should be interpreted faster than more challenging ones if we consider that the time lag corresponds to the latency response to a given task. However, the fact that the term is easily retrievable by the interpreter means that it burdens working memory less than a new term does, therefore the interpreter might feel more comfortable storing it for longer, while he/she might want to 'get rid of' the more challenging term as fast as possible. It should be noted, however, that the average EVS for challenging units was calculated using a mere four measurements, because the remaining five were omitted in the interpretation. As to mean EVS for *numbers*, only one measurement could be taken into account because the other two were not interpreted – it amounts to 8.13s. Since it has been shown that numbers decay quickly in working memory (Mazza, 2001), the tendency to keep a long EVS might be the reason why two out of three numbers were omitted.

Table 1: Data for performance A1

| category | tot cues | valid counts | omissions | average EVS (s) | SD (s) | counts excluding outliers | average EVS excluding outliers (s) |
|---|---|---|---|---|---|---|---|
| beginnings of clauses | 82 | 47 | 35 | 5,48 | 2,56 | 45 | 5,17 |
| keywords | 57 | 35 | 22 | 6,63 | 2,57 | 34 | 6,48 |
| challenging units | 9 | 4 | 5 | 5,90 | | | |
| numbers | 3 | 1 | 2 | 8,13 | | | |
| tot | 151 | 87 | 64 | | | | |

This performance presents a particularly high rate of omissions, representing 64 tags of the 151 that were applied to the input text. They stem mainly from a 50-second segment where the student slowed down and eventually stopped speaking. The difficulty of the speech may indeed have played a role in this. It should be worth mentioning that this breakdown is preceded by time lags higher than average, namely two word EVSs measuring 10.82 and 11.78 seconds – two of the higher values measured in this interpretation, the latter being an outlier. This occurrence is in line with the claim that a long EVS entails a higher risk of loss of content (Lamberger-Felber, 2001; Lee, 2002).

The long pause in the output of student A is the result of a breakdown – an overload probably resulting of a lack of experience in this kind of multitasking, especially when the difficulty of the speech is above a beginner's expected abilities.

> Undoubtedly, interpreters in live SI are certainly aware that pauses in their SI give a negative impression to an audience listening to the SI. In this situation, the longer pauses by interpreters in live SI mean that they are facing unavoidable obstacles in their information processing. (Lee, 2006:208)

In this specific case, novice interpreting trainees are unlikely to be preoccupied with the audience yet, however, they were indeed required to keep interpreting no matter the circumstances, therefore they would not have stopped unless they were uncapable of uttering an interpretation. As Davidson (2009:3) states:

> … students will occasionally lag further behind (mainly when they have failed to comprehend a segment, or when the speed of speech is too fast) and they will catch up only by discarding, uninterpreted, the segment or segments they either forgot or did not comprehend.

In this case, several segments of the speech were discarded due to overload and/or possibly because the student was overwhelmed by the speech which, as mentioned earlier, was aimed at students with a higher level of experience.

Aside from the two longer time lags, which were excluded as they were outliers, the two smaller values also stand out from the rest because they are the only ones below two seconds. The shortest comes after a six-second pause where the student only listened to the input, and the other one is the very first EVS, indicating the lag between the beginning of the speech and the beginning of the interpretation. The results in Lee (2002) showed that EVS in segments where the interpreter only listened to the input was shorter than in parts where the listening and speaking tasks overlapped, and here the case might be that the two values are smaller because the student was not speaking before the start of the segment concerned. These short time lags may indicate the aspiration to keep a short EVS that is however made unsustainable for a novice interpreter once speaking and listening begin to overlap.

### 5.1.2 Performance B1

Performance B1 has an average EVS of 3.61s (SD=1.45), obtained through 69 valid results at the *beginning of clauses*. After excluding two outliers, the resulting average EVS is 3.49s. The approach to EVS seems to be quite consistent, since *keyword* EVS, the mean of 45 valid results, is 3.70s (when excluding one outlier), not too dissimilar from the EVS for *challenging units*, which is 3.66s. In this case, too, few are the values which could be used to calculate the latter – three out of nine possible measurements. Twelve input keywords were not interpreted in B1, while 22 keywords were omitted in A1. However, this does not necessarily equal a loss of content. Some of these omissions can be ascribed to the fact that many of the keywords in the input were the word ‘drone’ (or ‘drones’), which is the main topic of the speech. The interpreter might therefore decide to imply it where it may sound redundant, especially since the Italian syntax allows for subjects to be implied. In this interpretation, all three *numbers* were interpreted, leading to an average EVS for this category of 5.06s – noticeably longer than the other average time lags.

Table 2: Data for performance B1

| category | tot cues | valid counts | omissions | average EVS (s) | SD (s) | counts excluding outliers | average EVS excluding outliers (s) |
|---|---|---|---|---|---|---|---|
| beginnings of clauses | 82 | 69 | 13 | 3.61 | 1.45 | 67 | 3.49 |
| keywords | 57 | 45 | 12 | 3.81 | 1.58 | 44 | 3.70 |
| challenging units | 9 | 3 | 6 | 3.66 | | | |
| numbers | 3 | 3 | 0 | 5.06 | | | |
| tot | 151 | 120 | 31 | | | | |

In this interpretation, 120 valid EVS values could be gathered. The two longest time lags recorded are word EVSs and they are the only values over eight seconds (8.58 for a keyword outlier and 8.26 for a number EVS). The keyword itself seems to be the reason for a slowed-down processing, since it is not preceded by long time lags, and the student pauses for a few seconds before being able to utter a translation of the keyword. This item is followed by a few omissions. The long number EVS instead appears to be influenced by increasingly long time lags, and is equally followed by omissions.

General EVS ranges from 1.25 to 7.77s. Both EVS values that were so long as to be considered outliers, measuring 7.77 and 7.18 seconds, were registered in similar circumstances, albeit at two separate points of the recordings. Both times, the clause which was interpreted using the outlier EVS was preceded by another clause whose EVS was longer than the clause itself. In other words, the student had yet to interpret the first clause (because he/she had stayed silent for a couple of seconds), when the speaker started uttering the following one. It is likely that the effort of keeping in mind both the finished clause and the new, incoming one slowed down the process and the utterance of the output.

In this performance, values below two seconds were recorded for 15 different cues (nine beginnings of clauses, five keywords, and one challenging unit). The interpreter uses short EVS at different times in the performance, but particularly around minute one, where, together with three of the time lags below two seconds, other time lags just above two seconds were recorded. At other points of the interpretation, short time lags followed a few seconds where the listening and the speaking tasks did not overlap.

## 5.2 Part two

### 5.2.1 Performance A2

In recording A2, the average of the 61 valid *beginning-of-clause* EVS values is 5.30s (SD=2.62). Without the two outlier values, average EVS is 5.03s. Average *keyword* EVS, obtained through 43 valid measurements, is 5.50s; 5.35s when excluding the two outliers. The six *challenging units* for which the time lag could be measured average at 5.61s. As in A1, the highest average is *number* EVS, amounting to 6.50s and obtained through all three possible measurements.

Table 3: Data for performance A2

| category | tot cues | valid counts | omissions | average EVS (s) | SD (s) | counts excluding outliers | average EVS excluding outliers (s) |
|---|---|---|---|---|---|---|---|
| beginnings of clauses | 82 | 61 | 21 | 5.30 | 2.62 | 59 | 5.03 |
| keywords | 57 | 43 | 14 | 5.49 | 2.21 | 42 | 5.35 |
| challenging units | 9 | 6 | 3 | 5.61 | | | |
| numbers | 3 | 3 | 0 | 6.50 | | | |
| tot | 151 | 113 | 38 | | | | |

Apart from the last category, EVS appears to be rather homogeneous in this interpretation. However, the standard deviations which were measured for the two larger groups are high, and EVS measurements range from 0.25s to 14.22s. The longest EVS, an outlier, is mostly composed of silence in the output. Prior to the segment concerned, the student had missed the beginning of a new paragraph and probably had to listen closely to what came next to understand the shift in the discourse. The previous segment had a long EVS of 5.99s and included the interpretation of a number, a highly demanding element which probably required too much concentration for enough to be available to the listening task of the overlapping input segment. After the pause, the student had to reintegrate the beginning of the new paragraph into the interpretation, otherwise the following information would have been out of context. This explains why, contrary to A1, in this case a period of silence was not followed by a short EVS. The student then tried to slowly catch up, shortening the lag from 14 to 12, to 10 seconds, but he/she missed three cues on the way and still had too long an EVS to hear the beginning of another new paragraph. This resulted in another silence in the output (about five seconds), but instead of trying to restore the beginning of the paragraph, this time the student let go of it, losing content, but managing to start another sentence with a shorter lag (3.42s). A shorter lag is not a guarantee of completeness, however: another silence follows shortly after, with the subject missing a few other cues.

Four of the five values over 10s are all concentrated in the portion of speech described above, where the student paused and missed content. The use of a long EVS therefore does not appear to be deliberate in this performance, although there is one occurrence where a 10.26s EVS was effective. At the beginning of the speech, the student encoded two clauses in a different order to the original, which resulted in one long and one very short EVS (below one second). This pairing paid dividends, because the message was conveyed without omitting any relevant information.

> Input: Until a few years ago, the average man and woman in the street when hearing the word "drone" **would probably think of an insect**. In the animal realm, <u>"drones" is the term used for male insects</u>, be it male wasp or male bee, that unlike females do not have a stinger.
>
> Output: Fino a qualche anno fa, la l'individuo medio per strada quando sentiva la parola "drone" <u>che in inglese significa *drone*</u> [word pronounced in an English accent] <u>drone e fuco</u>, **poteva pensare più che altro alla parola fuco**, che si riferisce al maschio delle api che non possiede un pungiglione.

In this instance, the student was able to use both types of EVS to his/her advantage, demonstrating a better command of EVS management than in part I.

The shortest EVS could be considered as an anticipation even if the output still comes after the input, because the interpreter only hears a syllable before he/she starts to encode the message. The student could have remembered from the first interpretation that the term 'drone' needed a disambiguation because it can mean two different things in English, which then have two different translations in Italian. The speaker makes a link between male bees and UAVs (unmanned aerial vehicles) by virtue of the fact that they are both called ‘drone’, which does not translate well into Italian. This caught student A off-guard the first time, which makes it likely for him/her to remember the challenge and to develop a strategy to overcome it during preparation time.

### 5.2.2 Performance B2

In interpretation B2, average EVS was calculated between 70 valid time lags at the *beginning of clauses*, which gave the result of 4.72s (SD=2.09). Without the two outliers, that average becomes 4.55s. Valid *keyword* EVS measurements are in the number of 47 and their average is 5.05s, which turns into 4.84s if we do not consider the two outliers. Five *challenging units* were interpreted with an average EVS of 3.98s. Finally, all three *numbers* were interpreted with an average EVS of 5.49s. Overall, the performance resulted in 26 null EVS measurements.

Table 4: Data for performance B2

| category | tot cues | valid counts | omissions | average EVS (s) | SD (s) | counts excluding outliers | average EVS excluding outliers (s) |
|---|---|---|---|---|---|---|---|
| beginnings of clauses | 82 | 70 | 12 | 4.72 | 2.09 | 68 | 4.55 |
| keywords | 57 | 47 | 10 | 5.05 | 1.99 | 45 | 4.84 |
| challenging units | 9 | 5 | 4 | 3.98 | | | |
| numbers | 3 | 3 | 0 | 5.49 | | | |
| tot | 151 | 125 | 26 | | | | |

EVS values span from 1.05 to 10.67s. Three values overall (two clause beginnings and one keyword EVS) exceed 10s and are considered as outliers, while seven EVSs (four beginnings of clauses, two keywords and one number) are shorter than two seconds. These small values are rather evenly distributed across the speech, there are no clusters of short time lags. In one instance, the segment interpreted with a short EVS comes after a two-second pause in the *input* which allowed the student to finish interpreting the previous segment before he/she had to start listening for what was coming next. Once again, this occurrence is in line with Lee (2002)’s findings whereby time lag is shorter when the listening and speaking tasks do not overlap.

However, after having started to interpret the segment with a short EVS, the student paused mid-sentence for four to five seconds, either because he/she could not find a word, or because he/she should have waited longer to gather more information. In another instance, the speech comprehension and speech production tasks did not overlap before the short EVS because the student had stopped speaking for three to four seconds, hence being ready to encode the input as soon as he/she heard it.

It may be worth spending two words on the short *number* EVS, since it seems to be evidence of a strategy adopted to approach numbers. Indeed, in this instance the student encoded the number almost as soon as he/she heard it and only afterwards did he/she encode the context to it, inverting the original order.

> Input: …to increase the theoretical survival chance <u>of anyone with a heart attack</u> **by ten times**.
>
> Output: …per incrementare le probabilità di sopravvivenza **di dieci volte** <u>in caso di attacco di cuore</u>.

As to the longer time lags, two of them concern the same segment, where the student spent between four and five seconds exclusively listening to the input, therefore lengthening the EVS which was already around six seconds. During those time lags, the student misses a challenging unit. The second highest value (a beginning-of-clause EVS) and the second keyword outlier (just below 10s) come shortly after another silence, the longest recorded in this performance. EVS values from segments close to it also are at the higher end of the spectrum, and one segment was omitted. As mentioned by Lee (2002, 2011), a long EVS affects not only the segment that is being interpreted, but also the following segment, with effects on its EVS length and its processing, potentially entailing a loss of content.

# 6. RESULTS

## 6.1 Comparison

### 6.1.1 Student A's performances

A remarkable difference between A1 and A2 is the amount of valid measurements, which increased by 26 in the second performance. Every category 'gained' in valid measurements: 14 more beginnings of clauses, eight more keywords, two more challenging units and two more numbers were interpreted.

Coincidentally, all categories had a shorter EVS in A2, which seems therefore to work in favour of a more complete interpretation. Some changes were contained – a negligible shift by 0.14s in general EVS and 0.29s for challenging units – while others were more noticeable – a decrease by 1.13s for keyword EVS and by 1.63 for number EVS. It should be noted, however, that the latter value is hardly significant, since only one number was interpreted in A1. Regardless, the student was able to interpret all three numbers in the second performance with a shorter lag, while in the first interpretation the tendency for a long EVS was probably to blame for missing two numbers. This could suggest that the student has been learning to encode highly demanding elements such as numbers more quickly, as advised by handbooks (e.g. Setton & Dawrant, 2016).

Table 5: intersection values for A1

| category | tot cues | average EVS (s) | SD (s) | counts excluding outliers | average EVS excluding outliers (s) |
|---|---|---|---|---|---|
| beginnings of clauses | 41 | 5.44 | 2.56 | 40 | 5.23 |
| keywords | 29 | 6.36 | 2.37 | 28 | 6.18 |
| challenging units | 4 | 5.90 | | | |
| numbers | 1 | 8.13 | | | |
| tot | 75 | | | | |

Because of the large gap in valid measurements between A1 and A2, it appeared necessary to add a comparison between only those items which were interpreted in both parts of the experiment, the ‘intersection’.

Table 6: intersection values for A2

| category | tot cues | average EVS (s) | SD (s) | counts excluding outliers | average EVS excluding outliers (s) |
|---|---|---|---|---|---|
| beginnings of clauses | 41 | 5.06 | 2.58 | 38 | 4.54 |
| keywords | 29 | 5.21 | 2.33 | 28 | 4.98 |
| challenging units | 4 | 7.25 | | | |
| numbers | 1 | 7.70 | | | |
| tot | 75 | | | | |

This comparison confirmed a general shortening of EVS. The decrease in general EVS is more noticeable in the intersection (0.69s instead of 0.14) and so is the decrease in keyword EVS, although to a more limited extent (1.20s instead of 1.13). On the contrary, number EVS decreases by a smaller margin (0.43s instead of 1.63). That is because the only element that was interpreted in A1 was the one with the longest EVS among numbers in A2. It could be that that specific number was an easy one for the student to store in his/her working memory, since he/she was able to interpret it with a long EVS even in A1, where long EVS led to the omission of the other numbers.

The only pattern that is reversed when we only consider values that were valid in both A1 and A2 is challenging-unit EVS, which increases by 1.35s. Indeed, the two shorter values which contribute to lowering average EVS for this category in A2 are the two that were not interpreted in A1, leaving only three common measurements. The different trends for keyword EVS and challenging-unit EVS shown by the intersection data would seem to reflect the effect of cognitive load. Keywords are more easily activated, therefore they are encoded more quickly than challenging units, which arguably need more time to be processed. This might be an indication that cognitive load still plays an important role in determining EVS at this point in the training of student A. However, as with number EVS, the averages for challenging unit EVS were calculated on the basis of few valid measurements, which makes this comparison less significant.

Moreover, in part II student A limited the overlapping of EVS segments, where working memory is more burdened because more than one clause has to be stored at the same time. The student has developed a tendency to avoid building up EVS length, particularly by letting go of information he/she has not been able to encode rapidly, possibly so that he/she does not miss the following segment.

Overall, EVS values cover a wide range in both part I and part II (SDs over two seconds), but the student appears to have gained a better ability to manage this variety of time lags in part II, as shown by the decreased number of breakdowns.

Table 7: comparison of average EVS in student A's performances, having excluded all outliers. Expressed in seconds.

| | A1 | | A2 | | A2-A1 | |
|---|---|---|---|---|---|---|
| category | speech average | intersection average | speech average | intersection average | speech difference | intersection difference |
| beginnings of clauses | 5.17 | 5.23 | 5.03 | 4.54 | -0.14 | -0.69 |
| keywords | 6.48 | 6.18 | 5.35 | 4.98 | -1.13 | -1.20 |
| challenging units | 5.90 | 5.90 | 5.61 | 7.25 | -0.29 | +1.35 |
| numbers | 8.13 | 8.13 | 6.50 | 7.70 | -1.63 | -0.43 |

### 6.1.1 Student B's performances

Student B's performances followed the opposite pattern overall: EVS values increased between part I and part II of the experiment. General average EVS expanded by 1.06s, keyword EVS by 1.14s, challenging-unit EVS increased by 0.32s, and number EVS was 0.44s longer. EVS appears to be much more varied across categories in B2 than it is in B1, where all averages except number EVS were between 3.61 and 3.81s. This result may suggest that the student takes a different approach to different categories of input items. EVS is also more varied within categories. Indeed, standard deviation is higher in B2 than it is in B1 for both general and keyword EVS, whether we consider the average of all values or only that of the intersection between B1 and B2. An increase in dispersion such as the one shown by SD may imply that the student is less restricted by the boundaries of cognitive capacity and is able to apply different time lags to different inputs.

In this case, comparing values from the intersection is less relevant because there is a much smaller difference in the number of omissions in B1 and B2 than there is between A1 and A2. Indeed, after such a comparison the results change only slightly, and all trends are confirmed. For this reason, tables showing intersection values for B performances are not provided in this chapter and can be found in appendix 6.

Table 8: comparison of average EVS in student B's performances, having excluded all outliers. Expressed in seconds.

| | B1 | | B2 | | B2-B1 | |
|---|---|---|---|---|---|---|
| category | speech average | intersection average | speech average | intersection average | speech difference | intersection difference |
| beginnings of clauses | 3.49 | 3.42 | 4.55 | 4.56 | +1.06 | +1.14 |
| keywords | 3.70 | 3.74 | 4.84 | 4.91 | +1.14 | +1.17 |
| challenging units | 3.66 | 3.66 | 3.98 | 4.01 | +0.32 | +0.35 |
| numbers | 5.06 | 5.06 | 5.49 | 5.49 | +0.43 | +0.43 |

Through the collected data, we can notice that number and challenging-unit averages, which were expected to decrease, increased instead, but to a more limited extent than the other categories. This may indicate the will not to store demanding elements for too long. One case of a short number EVS in B2 shows the student adapting the order of other sentence constituents to the position of the number so that it can be encoded quickly. Although this strategy is not applied throughout, it shows that the student is able to use it.

Keyword and general EVS increased substantially, possibly suggesting that the student has developed his/her WM capacity and may be more in control of EVS, deciding to wait for more information before uttering the interpretation (Davidson, 1992). The decrease in the use of short time lags (below two seconds) equally seems to support this.

Student B already seemed rather in control of the process in part I, managing to miss few of the set cues. A different approach consisting of a longer EVS which was more differentiated among categories equally appeared to be effective in part II. Contrary to expectations, average EVS fell within the range of the average found in existing literature (two to four seconds) in B1, while it was longer in B2.

## 6.2 Summary and Discussion

The results of the experiment show changes in the EVS of both students. Such changes follow different trends, with student A's EVS shortening and student B's lengthening. The former pattern seemed to be dictated by an effort to avoid unsustainable EVS lengths which caused breakdowns in the first interpretation. Over the following months of training, he/she could learn to not only keep a shorter EVS, but also differentiate the treatment of different situations. Student B's pattern appeared to stem from a preference for collecting more information before starting to encode the message. While this meant that student B's EVS, which fell within the range of the literature average (two to four seconds) in part I, exceeded it in part II, EVS in A2 did approach that range, as expected.

Contrasting trends were equally recorded for keyword EVS, which increased in B performances as expected, but decreased in A performances. Challenging-unit EVS, which we expected to decrease, increased in B2, but to a significantly lesser extent than the categories which were expected to increase. It did decrease in A2, when we consider all valid values in A1 and A2, but it increased according to the intersection values. Number EVS decreased as expected in A2 and increased in B2, but not as much as the categories we expected to increase. Number EVS, which was significantly higher than all other averages in B1, is still the longest in B2, but a specific instance of short number EVS in B2 denotes the use of a strategy to encode numbers quickly. EVS values are more varied in B2.

## 7. CONCLUSION

Opinions as to whether more experienced interpreters have longer or shorter EVS than novice interpreters vary in existing literature (Davidson, 1992; Moser-Mercer, 1997; Timarová et al., 2015; Ruiz Rosendo & Galvan, 2019). In this study, both trends were present: one student's time lag increased with experience and the other's decreased. Moreover, both approaches gave better results in terms of completeness of the interpretation, although the improvement is more noticeable for the student who decreased his/her EVS – student A, whose interpreted units count increased by 26, while student B's only increased by five. The students seem to be adapting their EVS to their respective requirements, and what could be considered their preferences in terms of EVS length start to appear. The hypothesis that EVS changes across time appears to be confirmed.

Word EVS did not differ as clearly as expected between categories. However, some trends were noticeable, namely challenging-unit EVS and number EVS which decreased, or increased proportionally less than keyword EVS and general EVS.

Nonetheless, the results did not show patterns clear enough, and there were not sufficient valid values to confirm or disprove the hypothesis concerning the different treatment of different word categories. More defined trends may emerge in a similar, bigger-scale experiment. The participation of more students and the interpretation of a custom-made speech with more elements to each category would mean that a bigger pool of data is collected, making changes and differences stand out more clearly if they do exist.

Moreover, changes may be more noticeable if the second part of the experiment is conducted further along the subjects' training. Indeed, some of the patterns observed suggest that the students have started to apply specific approaches to EVS and that they may apply them more consistently after further training. At this point in their training, they may still not be completely at ease performing SI, which would mean that they are still subjected to cognitive load more than they manage it, so they are not able to apply specific strategies as much as they may at the end of their training programme.

This study has shown that a within-subject design on student interpreters can reveal more about the adjustment of EVS to the requirements of the input and of the interpreter. The use of one and the same speech for both parts of the experiment clearly interfered with the results on one occasion and may have had other less noticeable consequences. A potential study using *two* custom-made speeches with the same characteristics would completely rule out this practice effect.

With the suggested alterations, a similar study could allow to make more accurate inferences on the evolution of EVS management.

## 8. REFERENCES


Adamowicz, A. (1989). The Role of Anticipation in Discourse: Text Processing in Simultaneous Interpreting. Polish Psychological Bulletin 20, 153–160.

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

American Psychological Association (2018). APA Dictionary of Psychology, Dual-Task Performance. Retrieved from https://dictionary.apa.org/dual-task-performance

American Psychological Association (2018). APA Dictionary of Psychology, Practice Effect. Retrieved from https://dictionary.apa.org/practice-effect

American Psychological Association (2018). APA Dictionary of Psychology, Secondary Task Methodology. Retrieved from https://dictionary.apa.org/secondary-task-methodology

Anderson, L. (1994). Simultaneous Interpretation: Contextual and Translational Aspects. Bridging the Gap: Empirical Research in Simultaneous Interpretation. Benjamins Translation Library 3, edited by Sylvie Lambert and Barbara Moser-Mercer, 101-120.

Baddeley, A. D. (2003). Working Memory: Looking Back and Looking Forward. Neuroscience 4, 829-839.

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

Barik, H.C. (1973). Simultaneous Interpretation: Temporal and Quantitative Data. Language and Speech 16(3), 237-270.

Bevilacqua, L. (2009). The Position of the Verb in Germanic Languages and Simultaneous Interpretation. The Interpreters' Newsletter 14, 1-31.

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

Chmiel, A., Szarkowska, A., Koržinek, D., Lijewska, A., Dutka, L., Brocki, Ł., Marasek, K. (2017). Ear-Voice Span and Pauses in Intra- and Interlingual Respeaking: An Exploratory Study into Temporal Aspects of the Respeaking Process. Applied Psycholinguistics 38(5), 1201-1227.

Christoffels, I. K., De Groot, A. M. B. (2004). Components of Simultaneous Interpreting: Comparing Interpreting with Shadowing and Paraphrasing. Bilingualism: Language and Cognition 7(3), 227-240.

Collard, C., Defrancq, B. (2019). Predictors of Ear-Voice Span, a Corpus-Based Study with Special Reference to Sex. Perspectives – Studies in Translation Theory and Practice 27(3), 431-454.

Davidson, P. M. (1992). Segmentation of Japanese Source Language Discourse in Simultaneous Interpretation. The Interpreter's Newsletter, Special Issue 1, 2-11.

De Groot, A. M. B. (1997). The Cognitive Study of Translation and Interpretation. J. H. Danks, G. M. Shreve, G.M. Fountain and M.K. McBeath (Eds.), Cognitive Processes in Translation and Interpreting, London: Sage Publications, 25-56.

Defrancq, B. (2015). Corpus-based Research into the Presumed Effects of Short EVS. Interpreting: International Journal of Research & Practice in Interpreting 17(1), 26-45.

Díaz-Gálaz, S. (2011). The Effect of Previous Preparation in Simultaneous Interpreting: Preliminary Results. Across Languages and Cultures 12(2), 173-191.

Diriker, E. (2015a). Conference Interpreting. Pöchhacker, F., Grbic, N., Mead, P. & Setton, R. Routledge Encyclopedia of Interpreting Studies. Oxford and New York: Routledge, 78-82.

Diriker, E. (2015b). Simultaneous Interpreting. Pöchhacker, F., Grbic, N., Mead, P. & Setton, R. Routledge Encyclopedia of Interpreting Studies. Oxford and New York: Routledge, 382-385.

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.

European Commission (n.d). How Does It Work? Which Speech Level Is Most Adequate for My Learning Stage? Retrieved from https://webgate.ec.europa.eu/sr/content/how-does-it-work-which-speech-level-most-adequate-my-learning-stage

Færch, C., Kasper, G. (1984). Two Ways of Defining Communication Strategies. Language Learning 3(1), 45-63.

Fusco, M. A. (1990). Quality in Conference Interpreting between Cognate Languages: A Preliminary Approach to the Spanish-Italian Case. The Interpreters' Newsletter 3, 93-97.

Gerver, D. (1969). The Effects of Source Language Presentation Rate on the Performance of Simultaneous Conference Interpreters. The Interpreting Studies Reader, 52-66.

Gerver, D. (1970). Aspects of Simultaneous Interpretation and Human Information Processing. Unpublished PhD thesis. Oxford, University of Oxford.

Gile, D. (1984). Des difficultés de la transmission informationnelle en interprétation simultanée. Babel 28, 18-25.

Gile, D. (1985). Les termes techniques en interprétation simultanée. Meta 30(3), 199-210.

Gile, D. (2015). Effort Models. Pöchhacker, F., Grbic, N., Mead, P. & Setton, R. Routledge Encyclopedia of Interpreting Studies. Oxford and New York: Routledge 135-137.

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

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

Haapalainen, E., Kim, S. J., Forlizzi, J. F., Dey, A. K. (2010). Psycho-Physiological Measures for Assessing Cognitive Load. Proceedings of the 12th International Conference on Ubiquitous Computing, 301-310. Copenhagen.

Healy, A. F., Miller, G. A. (1970). The Verb as the Main Determinant of Sentence Meaning. Psychonomic Science 20(6), 372.

Kim, H.-R. (2005). Linguistic Characteristics and Interpretation Strategy Based on EVS Analysis of Korean-Chinese, Korean-Japanese Interpretation. Meta 50(4).

Kohn, K., Kalina, S. (1996). The Strategic Dimension of Interpreting. Meta 41, 118-138.

Köpke, B., Signorelli, T. M. (2012). Methodological Aspects of Working Memory Assessment in Simultaneous Interpreters. International Journal of Bilingualism 16(2), 183-197.

Lamberger-Felber, H. (2001). Text-Oriented Research into Interpreting Examples form a Case-Study. Hermes, Journal of Linguistics 26.

Lambert, S. (1992). Shadowing. Meta 37(2), 263-273.

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

Lee, T.-H. (2003). Tail-to-Tail Span: A New Variable in Conference Interpreting Research. International Journal of Interpretation and Translation 1(1), 41-62.

Lee, T.-H. (2006). English into Korean Simultaneous Interpretation of Academy Awards Ceremony Through Open Captions on TV. Meta 56(1).

Lee, T.-H. (2011). A Comparison of Simultaneous Interpretation and Delayed Simultaneous Interpretation from English into Korean. Meta 51(2).

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

Miller, G. A. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information. Psychological Review 63, 81-97.

Moser-Mercer, B. (1997). Beyond Curiosity: Can Interpreting Research Meet the Challenge? J.H. Danks, G.M. Shreve, G.M. Fountain and M.K. McBeath (Eds.), Cognitive Processes in Translation and Interpreting, London: Sage Publications, 176-195.

Oléron, P., Nanpon, H. (1965). Recherches sur la traduction simultanée. Journal de Psychologie Normale et Pathologique 62(1), 73-94.

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

Paas, F. G. W. C., van Merriënboer, J. J. G. (1993). The Efficiency of Instructional Conditions: An Approach to Combine Mental Effort and Performance Measures. Human Factors 35, 737-743.

Paas, F. G. W. C., Tuovinen, J. E., Tabbers, H. K., van Gerven, P. W. M. (2003). Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. Educational Psychologist 38(1), 63-71.

Paneth, E. (1957). An Investigation into Conference Interpreting (with Special Reference to the Training of the Interpreter). Unpublished master's thesis, University of London Institute of Education.

Ruiz Rosendo, L., Galvan, M. C. (2019). Coping with speed: An Experimental Study on Expert and Novice Interpreter Performance in the Simultaneous Interpreting of Scientific Discourse. Babel 65(1), 1-25.

Schweda-Nicholson, N. (1987). Linguistic and Extra-Linguistic Aspects of Simultaneous Interpretation. Applied Linguistics 8, 194.

Seeber, K. (2011). Cognitive Load in Simultaneous Interpreting: Existing Theories – New Models. Interpreting 13(2), 176-204.

Seeber, K. (2013). Cognitive Load in Simultaneous Interpreting: Measures and Methods. Target: International Journal of Translation Studies 25(1), 18-32.

Seeber, K. (2015). Cognitive load. Pöchhacker, F., Grbic, N., Mead, P. & Setton, R. Routledge Encyclopedia of Interpreting Studies. Oxford and New York: Routledge, 60-61.

Seeber, K. (2017). Simultaneous Interpreting into a B Language: Considerations for Trainers and Trainees. Zybatow, L.N., Stauder, A. & Ustaszewski, M. (eds). Translation Studies and Translation Practice: Proceedings of the 2nd International Translata Conference, 2014: part 1, Frankfurt am Main: Peter Lang, 321-328.

Seeber, K., Kerzel, D. (2012). Cognitive Load in Simultaneous Interpreting: Model Meets Data. International Journal of Bilingualism 16(2), 228-242.

Setton, R. (1999). Simultaneous Interpretation: A Cognitive-Pragmatic Analysis. Benjamins translation library 28. Amsterdam; Philadelphia: John Benjamins.

Setton, R. (2001). Translation Studies and Cognitive Science: Do We Need Each Other? CTIS Occasional Papers 1, 113-126.

Setton, R., Dawrant, A. (2016). Conference Interpreting: A Complete Course. Amsterdam/Philadelphia: John Benjamins.

Timarová, S., Cenková, I., Meylaerts, R. (2015). Simultaneous Interpreting and Working Memory Capacity. A. Ferreira & J. W. Schwieter (Eds.), Psycholinguistic and Cognitive Inquiries into Translation and Interpreting, 101-126. Amsterdam: Benjamins.

Timarová, S. (2015a). Time Lag. Pöchhacker, F., Grbic, N., Mead, P. & Setton, R. Routledge Encyclopedia of Interpreting Studies. Oxford and New York: Routledge, 418-420.

Timarová, S. (2015b). Working Memory. Pöchhacker, F., Grbic, N., Mead, P. & Setton, R. Routledge Encyclopedia of Interpreting Studies. Oxford and New York: Routledge, 443-446.

Timarová, S., Dragsted, B., Hansen I. G. (2011). Time Lag in Translation and Interpreting, a Methodological Exploration. C. Alvstad, A. Hild & E. Tiselius (Eds.), Methods and strategies of process research: Integrative approaches in translation studies, 121-146.

Timarová, S., Cenková, I., Meylaerts, R., Hertog, E., Szmalec, A., Duyck, W. (2014). Simultaneous Interpreting and Working Memory Executive control. Interpreting 16(2), 139-168.

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

Yin, B., Chen, F., Ruiz, N. and Ambikairajah, E. (2008). Speech-Based Cognitive Load Monitoring System. Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, 2041-2044. Las Vegas, Nevada.

# 9. APPENDICES

## 1. Invitation to take part to the experiment

Dear ----,

I am writing to ask if you would be available to take part in an experiment aimed at observing the simultaneous interpreting process. The experiment constitutes the empirical part of my MA thesis supervised by Prof. Seeber. It will take place on two separate days: one day of your convenience from February 27 to March 1, and one day of your convenience in the week of May 27-31, should you agree to participate. The experiment will take approximately 25 minutes each time. It will be carried out in one of the rooms available to interpreting students at UniMail, University of Geneva.

If you decide to take part in this study, you will be asked to interpret simultaneously an English speech of approximately 5 minutes into Italian.

Your part in this study is confidential within legal limits. FTI and I will protect your privacy and the confidentiality of the data gathered for this experiment. All records of your performance will be stored anonymously.

Participation is voluntary and you have the right to drop out or refuse to participate.

If you are willing to participate please suggest a day and time. I am available for any questions.

Thank you,

Sara Dore

## 2. Experiment checklist

- O Check that the booths are active
- O Welcome the students
- O Explanation (Six-minute speech. The speaker will say the title, then start straight away. Whatever happens, please keep going. Make sure you turn on the mic, try to limit unnecessary noise. Make a sign when you are ready.)
- O Keywords. Time their preparation (10 minutes)
- O Sound-check
- O Start activity
- O Listen
- O Stop activity
- O Take back list of words
- O Thank the students

## 3. List of keywords given to the students ten minutes prior to the interpretation

| |
|---|
| DRONES TO CATCH MOSQUITOES |
| Drone |
| Stinger |
| Unmanned Aerial Vehicle (UAV) |
| Remote control |
| Belligerent |
| Realtors |
| Listings |
| Bird's-eye view |
| Trap |
| Biological agent |
| Offspring |
| Inhospitable areas |
| Microsoft project Premonition |
| Disease outbreak |
| Malaria |
| Defibrillator |
| GPS coordinates |
| Line of sight |
| Legal hurdles |

## 4. Challenging units and Numbers

| Challenging units | Numbers |
|---|---|
| top gun pilots | 10 |
| propellers | 50 |
| breeding grounds | 500 million |
| map the spread | |
| genetic make-up | |
| theoretical survival chance | |
| heart attack | |

## 5. EVS values

All valid values are expressed in seconds. Omissions are indicated as “NA”.

| EVS: | category | A1 | B1 | A2 | B2 |
|---|---|---|---|---|---|
| 1 | beginning of clause | 1.64 | 1.75 | 1.55 | 2.57 |
| 2 | beginning of clause | 3.18 | 3.63 | 3.16 | 4.31 |
| 3 | beginning of clause | 2.40 | 2.40 | 5.42 | 3.66 |
| 4 | keyword | 7.84 | 3.88 | 6.19 | 5.69 |
| 5 | beginning of clause | 5.06 | 3.34 | 10.26 | 4.99 |
| 6 | beginning of clause | NA | 3.82 | NA | NA |
| 7 | beginning of clause | NA | 2.86 | 0.25 | 6.16 |
| 8 | keyword | 4.01 | 3.01 | 1.17 | 8.07 |
| 9 | beginning of clause | NA | 3.66 | 4.62 | NA |
| 10 | beginning of clause | NA | 2.75 | 6.49 | 5.54 |
| 11 | keyword | NA | 1.82 | 5.98 | 4.61 |
| 12 | beginning of clause | 8.96 | 2.73 | 5.64 | 4.64 |
| 13 | keyword | 8.85 | 4.49 | 5.67 | 4.52 |
| 14 | keyword | 9.67 | 5.28 | 9.20 | 5.63 |
| 15 | beginning of clause | NA | 5.36 | 7.32 | 4.62 |
| 16 | keyword | NA | NA | NA | NA |
| 17 | beginning of clause | 13.85 | 2.82 | 6.19 | 2.10 |
| 18 | beginning of clause | NA | 1.42 | NA | 1.52 |
| 19 | beginning of clause | NA | 2.29 | 3.01 | 2.37 |
| 20 | beginning of clause | 3.61 | 2.22 | 2.85 | 2.23 |
| 21 | beginning of clause | 6.17 | 1.75 | 2.18 | 2.74 |
| 22 | beginning of clause | 8.65 | 2.39 | 4.43 | 2.40 |
| 23 | keyword | 8.40 | 1.81 | 4.26 | 1.66 |
| 24 | beginning of clause | 4.27 | 4.53 | 8.11 | 4.03 |
| 25 | keyword | 4.80 | NA | 4.80 | 3.14 |
| 26 | beginning of clause | NA | NA | NA | 2.86 |
| 27 | beginning of clause | 3.88 | 1.69 | 3.86 | 3.08 |
| 28 | beginning of clause | 6.31 | 3.67 | 5.23 | 3.73 |
| 29 | beginning of clause | 3.65 | 3.34 | 3.22 | 4.13 |
| 30 | keyword | 2.78 | 3.09 | 2.92 | 4.43 |
| 31 | keyword | NA | 5.69 | NA | 3.06 |
| 32 | beginning of clause | 4.49 | NA | 3.79 | 3.65 |
| 33 | beginning of clause | 5.59 | 2.62 | 2.99 | 3.83 |
| 34 | keyword | 8.35 | 2.40 | 3.67 | 3.30 |
| 35 | keyword | NA | NA | 5.65 | NA |
| 36 | keyword | 8.71 | 2.78 | NA | 5.07 |
| 37 | beginning of clause | NA | NA | NA | NA |
| 38 | keyword | 11.24 | 8.58 | 5.78 | 6.15 |
| 39 | beginning of clause | NA | NA | 7.77 | 5.10 |
| 40 | keyword | 2.70 | NA | 6.24 | 3.66 |
| 41 | beginning of clause | 3.34 | 1.68 | 4.32 | 1.98 |
| 42 | keyword | NA | 1.59 | 4.47 | 2.18 |

| | | | | | |
|---|---|---|---|---|---|
| 43 | beginning of clause | NA | 4.91 | NA | 7.41 |
| 44 | beginning of clause | 4.32 | NA | 2.37 | NA |
| 45 | challenging unit | NA | NA | 2.77 | 3.40 |
| 46 | beginning of clause | 3.96 | 2.65 | 3.02 | 3.93 |
| 47 | keyword | 6.43 | 4.70 | 2.62 | 2.70 |
| 48 | beginning of clause | 4.90 | 1.83 | 2.48 | 4.76 |
| 49 | keyword | 7.66 | 3.27 | 3.61 | 5.59 |
| 50 | beginning of clause | NA | 4.63 | NA | 7.33 |
| 51 | beginning of clause | 6.28 | 3.07 | 2.09 | 5.52 |
| 52 | keyword | 7.57 | 3.33 | 4.06 | 6.63 |
| 53 | keyword | 10.11 | 4.04 | 4.15 | 6.72 |
| 54 | beginning of clause | NA | 4.70 | 5.50 | 10.67 |
| 55 | keyword | 3.55 | 3.55 | 4.59 | 10.20 |
| 56 | beginning of clause | NA | 3.16 | 3.96 | 7.34 |
| 57 | challenging unit | NA | NA | NA | NA |
| 58 | keyword | NA | 3.07 | 5.76 | 5.28 |
| 59 | beginning of clause | 1.00 | 2.88 | 6.68 | 5.85 |
| 60 | keyword | 3.98 | 3.74 | 6.20 | 8.70 |
| 61 | beginning of clause | 3.58 | 3.39 | NA | 4.79 |
| 62 | challenging unit | NA | NA | 1.89 | NA |
| 63 | keyword | NA | NA | NA | NA |
| 64 | beginning of clause | 3.31 | 1.84 | 2.89 | 3.94 |
| 65 | beginning of clause | 5.98 | 5.03 | 5.50 | 5.13 |
| 66 | keyword | 7.59 | NA | NA | NA |
| 67 | keyword | 5.82 | 5.03 | 5.34 | 4.94 |
| 68 | beginning of clause | NA | 4.25 | 4.95 | 5.35 |
| 69 | keyword | NA | NA | NA | NA |
| 70 | beginning of clause | 3.04 | 3.88 | 4.09 | 3.65 |
| 71 | keyword | 2.61 | NA | NA | 3.50 |
| 72 | challenging unit | 5.92 | NA | 6.21 | NA |
| 73 | keyword | 4.35 | NA | 5.92 | NA |
| 74 | beginning of clause | 5.64 | 1.25 | 7.12 | 1.05 |
| 75 | keyword | 7.26 | 1.48 | 7.54 | 1.97 |
| 76 | keyword | 6.92 | 2.74 | 8.54 | 2.31 |
| 77 | beginning of clause | NA | 2.44 | NA | 2.32 |
| 78 | beginning of clause | 5.11 | 2.12 | 6.35 | 3.02 |
| 79 | keyword | 5.89 | 2.39 | 7.24 | 3.35 |
| 80 | keyword | 10.82 | 2.09 | NA | 3.99 |
| 81 | beginning of clause | NA | 3.69 | 5.01 | 7.13 |
| 82 | keyword | 6.30 | 3.43 | NA | 5.25 |
| 83 | beginning of clause | 10.50 | 4.21 | 3.84 | 5.21 |
| 84 | keyword | NA | 3.82 | 5.84 | 4.84 |
| 85 | keyword | NA | 5.83 | 8.86 | 9.62 |
| 86 | keyword | 11.78 | NA | NA | NA |
| 87 | beginning of clause | NA | 4.09 | 9.45 | 10.28 |
| 88 | beginning of clause | NA | NA | NA | NA |

| | | | | | |
|---|---|---|---|---|---|
| 89 | beginning of clause | NA | 3.63 | 6.49 | 7.29 |
| 90 | keyword | NA | 3.41 | 7.36 | 7.18 |
| 91 | beginning of clause | NA | 1.73 | 5.81 | 6.24 |
| 92 | keyword | NA | 1.73 | 9.59 | 6.24 |
| 93 | keyword | NA | 3.04 | 4.64 | 5.58 |
| 94 | keyword | NA | 3.89 | 4.92 | 5.74 |
| 95 | beginning of clause | NA | 4.85 | 7.32 | 5.25 |
| 96 | keyword | NA | 3.36 | 6.60 | 3.65 |
| 97 | beginning of clause | NA | 3.90 | 7.61 | NA |
| 98 | keyword | NA | 4.00 | 7.67 | NA |
| 99 | beginning of clause | NA | NA | NA | 3.48 |
| 100 | challenging unit | NA | NA | NA | 4.47 |
| 101 | keyword | NA | NA | NA | 5.61 |
| 102 | beginning of clause | NA | 2.79 | 4.38 | 3.61 |
| 103 | beginning of clause | NA | 2.54 | 2.14 | 1.92 |
| 104 | keyword | NA | 2.53 | 2.43 | 2.11 |
| 105 | keyword | NA | 4.18 | NA | 3.80 |
| 106 | beginning of clause | NA | NA | NA | 6.37 |
| 107 | keyword | 3.73 | 3.47 | 2.94 | NA |
| 108 | beginning of clause | 3.84 | 2.55 | 4.40 | 3.73 |
| 109 | beginning of clause | 3.04 | 4.21 | 3.11 | 2.41 |
| 110 | keyword | 6.06 | 4.94 | 3.87 | 6.24 |
| 111 | beginning of clause | 6.52 | 7.77 | 5.99 | 6.71 |
| 112 | number | NA | 8.26 | 6.77 | 6.87 |
| 113 | beginning of clause | 4.50 | NA | NA | NA |
| 114 | beginning of clause | 5.85 | NA | 14.22 | 5.71 |
| 115 | keyword | 7.44 | 6.03 | 11.49 | 5.35 |
| 116 | beginning of clause | 10.74 | 4.31 | NA | 6.57 |
| 117 | keyword | NA | NA | NA | 5.64 |
| 118 | beginning of clause | NA | NA | 12.12 | 4.13 |
| 119 | beginning of clause | 9.05 | 2.96 | 10.51 | 2.90 |
| 120 | challenging unit | NA | 5.83 | NA | 2.85 |
| 121 | beginning of clause | 4.99 | 6.07 | NA | 4.73 |
| 122 | keyword | 4.71 | 6.56 | 2.80 | 4.40 |
| 123 | challenging unit | 3.48 | NA | 6.03 | NA |
| 124 | beginning of clause | 3.39 | 6.35 | NA | 3.12 |
| 125 | keyword | 8.33 | 3.02 | 8.93 | 2.54 |
| 126 | beginning of clause | NA | NA | NA | NA |
| 127 | beginning of clause | 7.19 | 3.34 | 7.12 | 3.14 |
| 128 | challenging unit | 7.07 | 3.17 | 8.03 | 3.87 |
| 129 | beginning of clause | 7.10 | 2.83 | 6.71 | NA |
| 130 | challenging unit | 7.12 | 1.98 | 8.72 | 5.31 |
| 131 | beginning of clause | NA | 2.40 | 8.29 | 2.13 |
| 132 | number | 8.13 | 2.01 | 7.70 | 1.56 |
| 133 | beginning of clause | NA | 2.66 | NA | 3.21 |
| 134 | beginning of clause | NA | 3.80 | NA | 5.12 |

| | | | | | |
|---|---|---|---|---|---|
| 135 | beginning of clause | 2.55 | 6.18 | 7.64 | NA |
| 136 | beginning of clause | 3.10 | 6.44 | 3.42 | 8.41 |
| 137 | keyword | 2.90 | 6.91 | 2.94 | 8.55 |
| 138 | beginning of clause | 5.02 | 4.69 | 4.50 | 7.91 |
| 139 | keyword | 4.50 | 3.47 | 3.35 | 6.45 |
| 140 | beginning of clause | 7.12 | 5.19 | 6.86 | 8.29 |
| 141 | number | NA | 4.90 | 5.03 | 8.05 |
| 142 | beginning of clause | 7.37 | 4.85 | 5.70 | 8.53 |
| 143 | beginning of clause | NA | 3.63 | NA | 7.86 |
| 144 | beginning of clause | NA | 4.64 | NA | NA |
| 145 | keyword | NA | 2.16 | NA | NA |
| 146 | beginning of clause | NA | 4.38 | 3.12 | 4.05 |
| 147 | keyword | NA | 4.87 | 5.53 | 5.94 |
| 148 | beginning of clause | 7.14 | 7.18 | NA | 5.37 |
| 149 | beginning of clause | 8.15 | 6.27 | 5.09 | 5.47 |
| 150 | keyword | 8.55 | 6.89 | 4.96 | 5.79 |
| 151 | beginning of clause | 7.99 | NA | 2.60 | NA |

## 6. Student B's Intersection Data

intersection values for B1

| category | tot cues | average EVS (s) | SD (s) | counts excluding outliers | average EVS excluding outliers (s) |
|---|---|---|---|---|---|
| beginnings of clauses | 63 | 3.55 | 1.47 | 61 | 3.42 |
| keywords | 42 | 3.85 | 1.62 | 41 | 3.74 |
| challenging units | 3 | 3.66 | | | |
| numbers | 3 | 5.06 | | | |
| tot | 111 | | | | |

intersection values for B1

| category | tot cues | average EVS (s) | SD (s) | counts excluding outliers | average EVS excluding outliers (s) |
|---|---|---|---|---|---|
| beginnings of clauses | 63 | 4.75 | 2.17 | 61 | 4.56 |
| keywords | 42 | 5.14 | 2.06 | 40 | 4.91 |
| challenging units | 3 | 4.01 | | | |
| numbers | 3 | 5.49 | | | |
| tot | 111 | | | | |