Charles University in Prague

Faculty of Arts – Institute of Translation Studies

Philology – Translation Studies

University of Leuven

Faculty of Arts

Linguistics and Literary Studies – Translation Studies

Šárka T i m a r o v á

Working memory in conference simultaneous interpreting

Doctoral dissertation

Supervisors: Prof. PhDr. Ivana Čeňková, CSc.
Prof. Reinhilde Meylaerts

Co-supervisor: Prof. Erik Hertog

2012

Acknowledgments

A doctoral dissertation is a written account of an individual's research project, which is not quite true really. This work would never have been possible without the help, support and advice of many people who have joined me on this journey.

I am extremely grateful to my supervisors, Ivana Čeňková, Erik Hertog and Reine Meylaerts for their support, advice and patience. I am equally grateful to Frieda Steurs and Lessius University College for financial and institutional support. I would like to thank Miriam Shlesinger and Ingrid Christoffels for reading and commenting on the initial proposal for this dissertation. The help and guidance of Arnaud Szmalec and Wouter Duyck was instrumental during the design of tests and development of testing materials. Dirk Hoefkens made my mobile testing laboratory possible and also made all video recordings for the interpreting tasks. The speeches were delivered by David Chan and Tony Scott. Ann Matthyssen helped me to develop all materials which needed to be in Dutch. Dirk Reunbrouck was my sounding board for many of the ideas that found their way to this study. Ana Castro Sotos gave valuable advice for statistical analyses. Franz Pöchhacker has always shown keen interest in my work and offered his wise advice. Finally, I would like to thank the twenty-eight anonymous interpreters for their enthusiasm and stamina.

I hereby declare that this submission is my own work, that the work of others has been duly acknowledged and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of a university or other institute of higher learning, except where due acknowledgement is made in the text.

Abstract

This dissertation addresses the question of working memory involvement in conference simultaneous interpreting. Working memory is considered to be a crucial cognitive mechanism for the simultaneous interpreting process, but this assumption has not been substantially supported empirically. The present study builds on an analysis of theoretical literature in both cognitive psychology and Interpreting Studies. A review of previously published empirical research identified several gaps, namely that research focused on a limited range of working memory functions (primarily memory functions) and on participants with no or limited interpreting experience. The design of the present study includes a wider range of working memory tests, with emphasis on central executive functions (related to attention) and on recruitment of professional interpreters. Interpreters' performance on working memory tests and an interpreting task was compared. The results show that a) interpreters' working memory is related to their performance in simultaneous interpreting, b) that simultaneous interpreting is predominantly related to the central executive functions, but not to memory functions, c) that some working memory functions seem to develop with interpreting experience, while others do not, and d) that the relationship between working memory and simultaneous interpreting is best described as many-to-many, i.e. different working memory functions predict different sub-processes in simultaneous interpreting in multiple complex patterns. The conclusions of this study are data-driven, but in line with current literature. More specifically, the findings support those accounts of simultaneous interpreting which emphasise attentional control as an important component of simultaneous interpreting processing.

Key words: simultaneous interpreting, working memory, central executive, attentional control

Contents

List of figures

List of tables

Introduction

The concept of working memory was proposed in the 1970s by Baddeley and Hitch (1974) as a modification of the concept of short-term memory. The crucial difference was that short-term memory was a simple store for information, while the more modern concept of working memory is assumed to be a cognitive component combining storage, processing and executive control of the cognitive processes at hand. Working memory is one of the most prominent topics in both current cognitive research and interpreting research. With more empirical findings available, working memory is now linked to a large number of higher-cognitive abilities and processes, and also to intelligence. In interpreting, it is considered to be one of the cognitive cornerstones underlying simultaneous interpreting (SI; Darò, 1989; Bajo, Padilla and Padilla, 2000). A number of studies focused on showing this by demonstrating a difference in working memory between simultaneous interpreters and SI students or non-interpreters. However, the results are mixed: some studies have found a difference (e.g. Padilla, Bajo, Cañas and Padilla, 1995; Christoffels, 2004), others have not (e.g. Chincotta and Underwood, 1998; Nordet and Voegtlin, 1998; Köpke and Nespoulous, 2006).

The goal of the present dissertation is to address the question of working memory in simultaneous interpreting (SI) in a broadly designed correlational study, which explores the relationships between individual working memory functions and various aspects of simultaneous interpreting performance. In Chapter 1, we describe three models of working memory proposed within cognitive psychology and by contrasting them, we highlight the complexity of working memory research. We then selected three models of the simultaneous interpreting process and discuss the role of working memory in simultaneous interpreting as it has been stipulated on a theoretical, model-building level. In Chapter 2, we focus on the role of working memory in general cognition. Drawing on cognitive psychological literature, we discuss the empirical relationship between working memory and language processing, bilingualism and skilled behaviour, three areas relevant in simultaneous interpreting. In Chapter 3, we turn our attention to previous empirical research relating working memory and simultaneous interpreting. Analysis of available literature shows that research has generally focused on the question of interpreters' larger working memory capacity in comparison to non-interpreters ('superiority hypothesis'), and that studies relating working memory to simultaneous interpreting performance, rather than interpreters, were mostly conducted with non-professional participants. This gap in previous research, i.e. lack of research into working memory and simultaneous interpreting performance in professional interpreters, then serves as a basis for formulating our own research questions at the end of Chapter 3. Given the relative paucity of previous research and the ambiguous findings reported by others, we opt for an exploratory correlational study. While previous research employed both “working memory” and “simultaneous interpreting performance” as unitary constructs, our aim is to emphasise the highly complex structures of both. To that end, both working memory and simultaneous interpreting performance are operationalised as a range of distinct functions (working memory) and subprocesses (simultaneous interpreting). This fine-tuned approach should allow us to study more specific relationships, potentially specifying which of the many working memory functions are involved in SI. Chapter 4 describes both the general methodological approach and the specific design, materials and procedure. In Chapter 5, we present the results and related discussion in three sections: structure of working memory in interpreters, structure of the SI task, and relationship between working memory and simultaneous interpreting. Finally, in Chapter 6 we discuss the more general implications and limitations of the study, and make suggestions for further research.

Chapter 1. Working memory[1]

1.1. The nature of working memory

Peter gets two apples from his mother every day, but only eats one. How many apples does he have at the end of the third day? This classic example of mental arithmetic is a good illustration of what working memory is for. To solve the problem, we need to make some stepwise calculations and store the result of each step in order to integrate it with the result of the next step. This type of memory required for the short-term storage of information was first described by William James (1890) as *primary memory*. Research in the 1950s and 1960s showed that the store has a severely limited capacity. The seminal article by Miller (1956) claims humans can at any one time remember a maximum of seven, plus or minus two, chunks. However, the simple storage account did not adequately explain a range of empirical data. This led to the proposal of *working memory*, which combined storage with ongoing processing (Atkinson and Shiffrin, 1971; Baddeley and Hitch, 1974). The concept of working memory in this form has become one of the key constructs in modern cognitive disciplines and through empirical research has been linked to a host of everyday activities, mental abilities and higher-cognitive processes. Working memory is thus known to be related to reading comprehension (Daneman and Carpenter, 1980), language comprehension (Just and Carpenter, 1992), reasoning (Kyllonen and Christal, 1990), vocabulary learning (Baddeley, Gathercole and Papagno,1998), problem solving (Kyllonen and Christal, 1990), or taking lecture notes (Kiewra and Benton, 1988). There is also a growing body of evidence that working memory is strongly related to general intelligence (Engle, Tuholski, Laughlin, and Conway, 1999; Conway, Cowan, Bunting, Therriault and Minkoff, 2002; Redick and Engle, 2006; Jarrold and Towse, 2006). Quite naturally, the huge amount of data collected gave rise to the proposal of numerous theoretical models.

1.2. Conceptualising working memory: models

Since the 1970s, at least a dozen theoretical models have been developed to account for all the findings. In this section, we will focus on only three of them. The selection has been guided by several factors: the model should be representative of the cognitive research, it should be current and updated, and it should provide theoretical background for our

---

[1] Chapter 1 is adopted from Timarová (2007) and was published as a CETRA paper (Timarová, 2008).

discussion of working memory research as it relates to the interpreting process. For this reason, the models briefly discussed in this section will be the multicomponent model by Baddeley (Baddeley and Hitch, 1974; Baddeley, 1996; Baddeley, 2000; Repovš and Baddeley, 2006); the long-term working memory model by Ericsson and Kintsch (1995); and the long-term memory activation model by Cowan (Cowan, 1988, 1995; Oberauer, 2002), and its control-of-attention extension by Conway and Engle (Conway and Engle, 1994; Engle, Tuholski, Laughlin and Conway, 1999; Feldman Barrett, Tugade and Engle, 2004).

1.2.1. Baddeley: A multicomponent model

One of the most influential working memory models to date was developed by Alan Baddeley (Baddeley and Hitch, 1974; Baddeley, 1996; Baddeley, 2000). Based on a wide range of empirical findings, this model posits that working memory is composed of separate storage and processing systems. The current version of the model is shown in Figure 1.1.

The model assumes that working memory is composed of two domain-specific slave storage systems - the *phonological loop* and the *visuospatial sketchpad*, a general storage component - the *episodic buffer*, and a supervisory component - the *executive control*. Each of the two slave systems is responsible for temporary storage of domain-specific information, i.e. the phonological loop stores verbal and numerical information, the visuospatial sketchpad specialises in the storage of visual and spatial information. Each slave system is further fractionated into subcomponents.[2] The phonological loop is thus composed of a phonological store and an articulatory control process. The phonological store allows for direct access to auditorily presented verbal information, while visually presented text must first be subvocally articulated by the articulatory control process before it can be stored in the phonological store.

[2] Since our concern lies mainly with interpretation, detailed discussion will be limited to the phonological loop. It is assumed that the visuospatial sketchpad plays no or only a marginal role in interpreting, a predominantly verbal task.

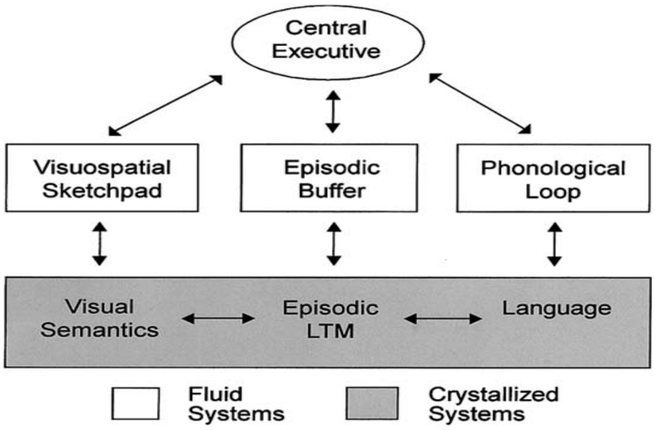

Figure 1.1 The multicomponent working memory model (From Baddeley, 2000)

Capacity of the working memory is defined as a temporal limitation on the amount of information that can be stored. Empirical findings by Baddeley et al. (1975) have shown that the size of the phonological store is approximately two seconds of verbal material, i.e. people can remember as many words as they are able to pronounce in two seconds. The two slave systems are dependent on a central executive, a supervisory component which controls and coordinates mental operations (Hitch, 2005).

Most research has been devoted to exploring the phonological loop (Baddeley, 1996), as its ability to account for a large range of findings proved intriguing. However, by the mid-1990s, Baddeley called for closer exploration of the central executive (Baddeley, 1996), a neglected, yet probably the most important of the working memory components. Further research and evaluation of the model led the author to add another slave system, the episodic buffer (Baddeley, 2000). This latest addition was motivated by the need to solve some persistent problems, such as the interface with long-term memory or the need to account for such phenomena as prose recall (recall of coherent strings of text such as complete sentences, as opposed to lists of individual words), where people typically score much higher than would be possible according to the limitations assumed for the phonological loop. The episodic buffer thus serves as a mental workspace which stores processed and integrated information. The model, as depicted in Figure 1.1, also allows for the distinction between fluid systems, which serve general processing (e.g. temporary storage), and crystallized systems, which allow for long-term storage and accumulation of knowledge (Baddeley, 2000). The fluid systems are supposed to be fairly stable, unaffected by learning, while the crystallized systems are the exact opposite – very much the result of learning. This distinction is important in the context of interpreting as an acquired skill. A large proportion of research into working memory in interpreting to date has focused on testing those components of working memory which are classified as fluid systems. Based on the theoretical model, and the predictions derived from it, it is so far not clear why interpreters should exhibit larger working memory capacity of the fluid components. This issue will be discussed further in Chapter 3.

1.2.2. Ericsson and Kintsch: Long-term working memory

A different approach to working memory was adopted by Ericsson and Kintsch. While most working memory research aims at discovering the nature of pure working memory, Ericsson and Kintsch are interested in how working memory functions in more realistic scenarios. Specifically, they address the question of how working memory supports skilled performance, ranging from instances of highly specialised activities, such as expert chess playing, to more everyday and yet very skilled activities such as reading and comprehension. Their analysis of such expert performance led to the proposal of long-term working memory (LT-WM; Ericsson and Kintsch, 1995). According to Ericsson and Kintsch, the traditional working memory (or short-term memory) concept is incompatible with real-life skilled performance on at least four counts: experts are capable of maintaining much more information than predicted by the capacity limits constraining WM; expert performance can be interrupted by another, attention-demanding task and later resumed with very little disruption to memory; experts are very accurate in their recall even if the recall task has not been expected; and the storage capacity of STM is fixed, while experts are capable of expanding it. All these issues are addressed in the LT-WM model. Crucially, Ericsson and Kintsch do not attempt to replace existing STM and WM models, but rather make an addition. Their LT-WM model arguably applies only to behaviour and tasks which are well practiced and to material which is familiar. LT-WM is proposed to be a "set of acquired mechanisms that enables experts to expand the functional capacity of their working memory system for specific types of materials in activities within their domain of expertise without altering the general capacity limits of [short-term working memory]" (Ericsson and Delaney, 1998:95). Importantly, LT-WM relies on storage in LTM. This is the main explanatory element behind the above mentioned issues: the amount of information can exceed the traditional STM capacity because it is stored in LTM and as such is more durable, hence less vulnerable to disruption. LT-WM is thus acquired, specific to a particular area of expertise, and the mechanisms supporting it are not directly transferable to another domain (for example, expert memory for digits does not imply expert memory for consonant strings - Ericsson and Delaney, 1998).

The above claim about LT-WM being an addition rather than a substitution needs qualification. The area of overlap between LT-WM and more traditional WM models becomes apparent when attempting to define an “area of expertise”. Ericsson and Kintsch (1995) include under expert performance such activities as reading and comprehension. These areas are in the focus of the traditional WM research as well, hence assuming that text which is being processed during reading and comprehension is being transiently stored in a separate WM store, while LT-WM assumes that reading is a skilled activity supported by LTM (Ericsson and Kintsch, 1995; Kintsch, 1998). Furthermore, the LT-WM concept does not do away with STM completely. While most information necessary for a successful execution of the task is available in LTM, it is accessed through specific retrieval cues available in STM. This seems to be the weaker point of the LT-WM model. The authors do not explain the mechanisms supporting the maintenance of the retrieval cues. Assuming that they are stored in the traditional STM (Ericsson and Delaney, 1998), they should be subject to the same limitations and constraints as other non-familiar material, such as a string of unrelated digits, i.e. there should be a capacity limit on the number of cues that can be maintained active, their maintenance should be vulnerable to interference, and the number of cues that can be recalled should be fairly fixed (once the number of cues reaches the capacity of the STM store, as the capacity is assumed to be stable throughout one’s adult life).

1.2.3. Cowan: Activated long-term memory

A third approach was adopted by Nelson Cowan (1988, 1995), who conceptualises working memory as an activated part of long-term memory (Figure 1.2). According to this view (Cowan, 1999), long-term memory contains an activated subset of items. These items are the content of working memory, but are outside awareness. Only a small proportion of these items are directly accessible to awareness: these items are, in Cowan’s words, in the focus of attention. The concept of working memory is thus hierarchical – the focus of attention contains a small number of items directly available for processing. These are a subset of the activated memory. The activated memory contains highly activated items which are, nonetheless, available for processing only indirectly through inclusion in the focus of attention. This activated memory is in turn a subset of long-term memory. Items in long-term memory are not active, but can become a part of working memory if they receive enough activation (and eventually they can move to the focus of attention). The crucial component, the focus of attention, is controlled by a central executive component responsible for directing attention and controlling voluntary processing. Moreover, attention can be captured more directly and automatically by stimuli (e.g. loud noise or stimuli perceived as relevant, such as one’s own name, etc.).

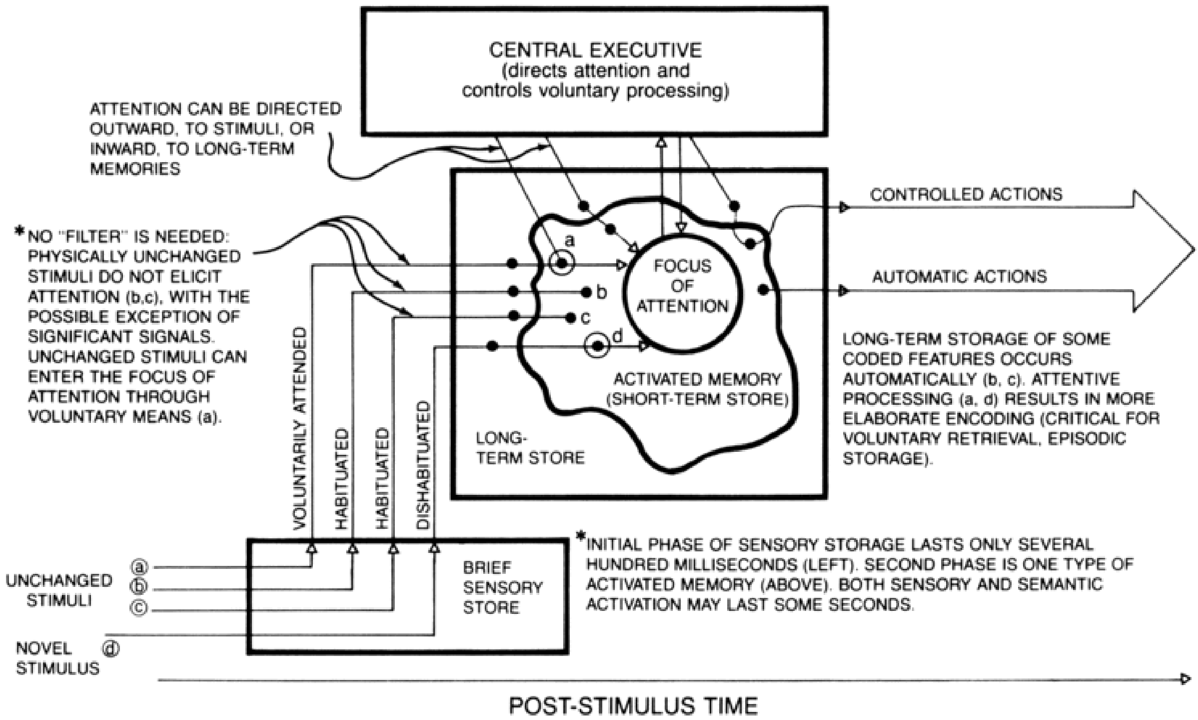

Figure 1.2 The working memory as an activated part of long-term memory (From Cowan, 1988)

Cowan proposes two types of limitations on working memory: time limitations and capacity limitations (Cowan, 1999). The focus of attention is subject to capacity limitations. Its estimated capacity is around ± 4 items. On the other hand, the activated memory is subject to time limitations. Unless activation of the items within the activated memory is maintained, it will decay and eventually drift out of the activated memory. The decay time is estimated to be around 10-30 seconds. Activation can be maintained through a rehearsal-like process of moving the item into the focus of attention. Interestingly, Cowan suggested that there is no capacity limit on the number of items concurrently activated, i.e. held in the activated memory.

The basic tenets of Cowan's model have been recently empirically supported and extended by Oberauer (Oberauer, 2002; Oberauer and Göthe, 2006; Oberauer, 2006). Oberauer's model builds on Cowan's model and develops further the mechanisms of the focus of attention. Moreover, work by Conway and Engle (e.g. Conway and Engle, 1994; Engle et al., 1999; Engle and Kane, 2004) further provides both theoretical argument and empirical support for Cowan's model.[3] Their claim is that the crucial component of individual differences in working memory is controlled attention, responsible for maintaining goals, protection of task execution against interference, effortful processes, etc. They extend

[3] However, Cowan sees some important differences between his work and the work by Engle and colleagues (Cowan, 1999).

Cowan’s basic model by explicitly including the executive component and by allowing for domain-specific strategies (see Figure 1.3).

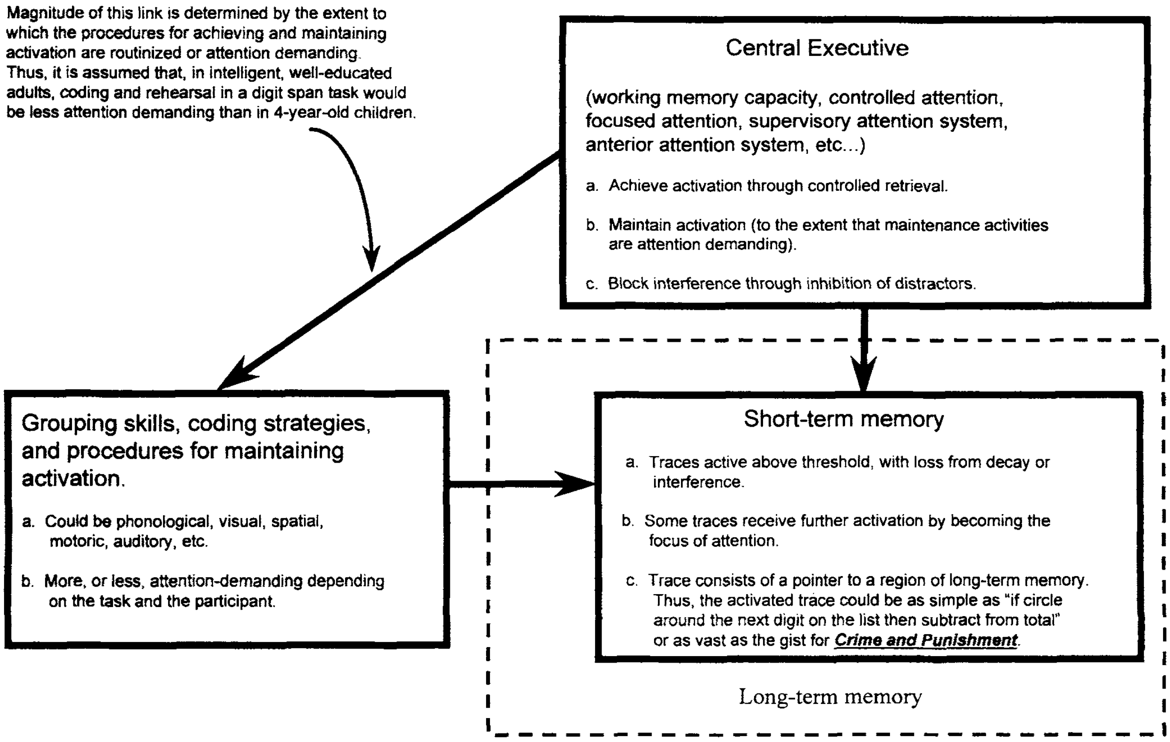

Figure 1.3 The working memory structure model by Engle, Tuholski, Laughlin and Conway (1999). (From Engle et al., 1999)

1.3. Working memory models: Discussion

In the previous section, we offered a short overview of three models of working memory. It should be clear by now that working memory as a concept is a very complex phenomenon. Each model places a different emphasis on different aspects of working memory, and reconciliation of the models is by no means easy. In this section, we will point out some of the major differences between the models.

The first difference concerns the view of working memory as a structural or a functional entity. Baddeley’s model is a good example of a structural model (cf. Engle and Kane, 2004). The basic assumption is that working memory constitutes a separate entity with more or less independent components, and the research aim is to identify the components and define their properties. On the other hand, Ericsson and Kintsch and Cowan postulate their models more in functional terms. This distinction is important for several reasons. The structural approach assumes that working memory is a separate entity with a fixed structure. The structure supports (and limits) cognitive performance. The functional approach, on the other hand, defines working memory in terms of processes or purpose, rather than structures. Working memory is thus defined as "temporary storage and processing of information" (functional view), and it can be supported by any neuroanatomical structure, including support by different structures on different occasions. Working memory is seen as a set of mechanisms which enable short-term maintenance of task-relevant information to ensure successful task completion (cf. Cowan, 1999). The models by Cowan, and Ericsson and Kintsch are less concerned with identifying specific "stores", but rather focus on the processes. Both Ericsson (Ericsson and Delaney, 1999) and Cowan (1999) openly refuse to limit their models to a predetermined structure, allowing working memory to make use of whichever mechanisms are available or appropriate. The reason for this difference may be partly historical. Baddeley's model evolved from earlier models assuming a necessary separate store for information which needs to be maintained for short periods of time. Ericsson and Kintsch's and Cowan's models are younger and address the more recent issue of interface between working memory (current task processing) and long-term memory (knowledge structures). The conceptual dissociation of working memory and long-term memory certainly holds, but there are doubts as to their neuroanatomical dissociation (e.g. Ranganath and Blumenfeld, 2005; Ruchkin, Grafman, Cameron and Berndt, 2003). Baddeley and Logie (1999) defend the separate-stores view, and point out the important function of WM as a gateway of novel stimuli, which were not previously encountered and hence have no representation in long-term memory.

Secondly, and more importantly for our topic, the models differ in their treatment of an executive component. Ericsson and Kintsch do not explicitly assume a separate executive component. Skilled performance seems to be guided mostly by processes relying on previous experience. Performers anticipate upcoming stimuli, use goal-directed strategies, rely on various cues, etc. On the other hand, Baddeley and Cowan assign a crucial role to their executive components. In both models, this component is very closely related to attention, and it is associated with effortful, conscious processing. In Cowan's model, the central executive controls and directs attention and voluntary processing. Similarly, Baddeley assumes that the central executive is responsible for the coordination of two (or more) tasks, selective attention, strategies switching, and manipulation of information in long-term memory (Baddeley, 1996; Baddeley, 2002). The crucial role of the central executive is supported by current theoretical research, which seems to converge on the idea that the central executive is largely synonymous with controlled attention. Controlled attention thus emerges as a separable entity. Moreover, individual differences research suggests that controlled attention is more important (is a better correlate) for higher-cognition activities than the storage component of working memory (e.g. Engle, Tuholski, Laughlin and Conway, 1999; Conway and Engle, 1994; Feldman Barrett, Tugade and Engle, 2004; Lépine, Bernardin and Barrouillet, 2005; Cowan, Elliott, Saults, Morey, Mattox, Hismajatullina and Conway, 2005; Engle, 2002; Hester and Garavan, 2005; but see Colom, Rebollo, Abad and Shih, 2006; Buehner, Krumm and Pick, 2005 for a view supporting the storage component as the most important element of working memory).

Thirdly, the models differ on the assumed capacity of working memory, both in terms of the size of the working memory, and in terms of the nature of its limitations. Baddeley assumes a temporal limitation, whereby the phonological loop can maintain approximately 2 seconds worth of verbal material. The actual amount of material stored can differ among individuals depending on their speed of articulation: those who can pronounce more in two seconds are also able to remember more (but see Baddeley and Logie, 1999, for a less definitive view on the working memory capacity). Cowan (1999), as mentioned above, assumes a double limitation, both temporal and in terms of capacity. On the one hand, items in the activated memory are subject to temporal decay within 10-30 seconds, if their activation is not maintained. On the other hand, items in the focus of attention are limited to ±4 items, but can be held in attention for a substantial amount of time (Cowan, 1999). Ericsson and Delaney (1999) assume that in principle there is no limit on the number of activated items. However, as mentioned above, Ericsson and Kintsch's model assumes retrieval cues to be maintained in short-term memory, and it is not clear how many cues can be held active at the same time and for how long.

To conclude, in this short overview we presented three models of working memory which we believe are representative of the current debate in the field, and which are at the same time relevant for interpreting research. We highlighted the issues of structural vs functional modelling, executive control and attention, and the nature and quantification of working memory capacity.

1.4. Working memory and interpreting

In the long tradition of research into cognitive processes involved in (simultaneous) interpreting, working memory is among the components that received most attention in both theoretical writing and empirical studies. Basically all major cognitive (process) models of SI assume that working memory plays a crucial role, and several models are built directly on the concept (Gerver, 1976; Moser, 1978; Darò and Fabbro, 1994). It must be mentioned here that the majority of theoretical considerations of working memory and its role in SI are limited to storage functions. Executive functions have rarely been taken into considerations, and to date no empirical testing of the central executive has been carried out. The aim of this section is to sketch out some of the lines of thought on the role of working memory in simultaneous interpreting. Similarly to the section on general models of working memory, this section is very selective in terms of the models to be discussed. The objective is to provide an analysis of the various approaches to working memory, rather than an exhaustive overview of all previous work on the topic.

1.4.1. Process models of simultaneous interpreting and working memory

One of the first cognitive models of simultaneous interpreting was introduced by David Gerver in the early 1970s (Gerver, 1975, 1976). Gerver conducted a number of experiments with professional interpreters and based on the results proposed a sequential model of mental processing during interpreting (Figure 1.4). The model focuses on a system of short-term stores for the different stages of text processing. Gerver assumed that the source text is stored in an input buffer, from where it proceeds for further processing. The input buffer also stores a segment of the input text while the processor is busy with a previous segment.

The actual text processing is, according to Gerver, performed in co-operation with long-term memory, which activates the appropriate linguistic units. Gerver assumed that this stage of processing is purely linguistic, and did not consider it any further (Gerver, 1976). The processed material is then ready for output – via an output buffer, where it undergoes optional monitoring.

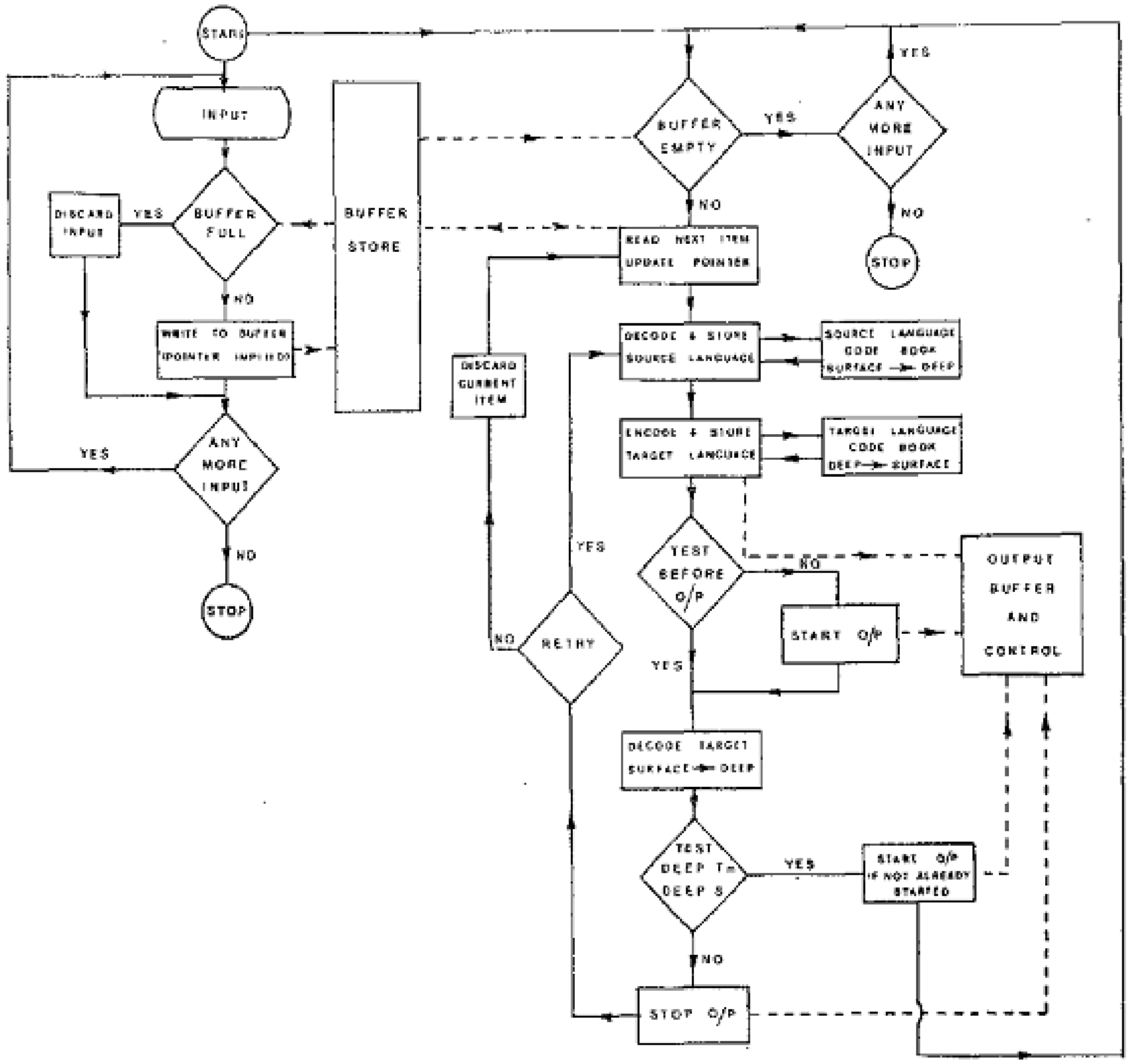

Figure 1.4 Model of simultaneous interpreting by David Gerver (1975)

Another process model of simultaneous interpreting was proposed by Barbara Moser in the mid-1970s (Moser, 1978; Figure 1.5). Her model too assigns a crucial role to working memory. Unlike Gerver, Moser is more explicit as to specific process segments. Barbara Moser uses a different term – *generated abstract memory* (GAM), but mentions specifically that it is identical to short-term memory (Moser, 1978). Her model is also interesting on a number of levels. First of all, Moser seems to consider working memory (at least as far as is apparent from the depicted model and from the accompanying verbal description) to be both a structural and functional component. GAM performs several important tasks. It stores processed chunks of text (strings of syntactically and semantically processed words, although it is not clear to which extent they are processed). This is the memory (storage) function proper. GAM also performs a recoding task in co-operation with a conceptual base, i.e. it is involved in the linguistic transformation. Throughout these operations, GAM is strongly linked to long-term memory, which stores all concepts, mental lexicons, syntax and grammar rules, etc. In Moser's view, GAM is explicitly involved in production as well. This makes the placement of the paraphrasing and prediction functions outside GAM somewhat incoherent. In the model, GAM does not encompass processes of auditory perception (at the beginning) and articulation (at the end). Thus the reasons for the exclusion of paraphrasing and predictions are not entirely clear, nor is it clear which other structural component performs these tasks. A second interesting aspect of this model is that GAM is equated with short-term memory, and that its storage function is emphasised. Short-term memory in cognitive psychology in the early 1970s was considered to be a purely passive store. Moser's model is ambiguous in this respect. By including the recoding functions in GAM, Moser proposes a very modern concept of working memory, which seems to include executive functions. Unambiguous interpretation of her model is, unfortunately, impossible on the basis of her short article (Moser, 1978).

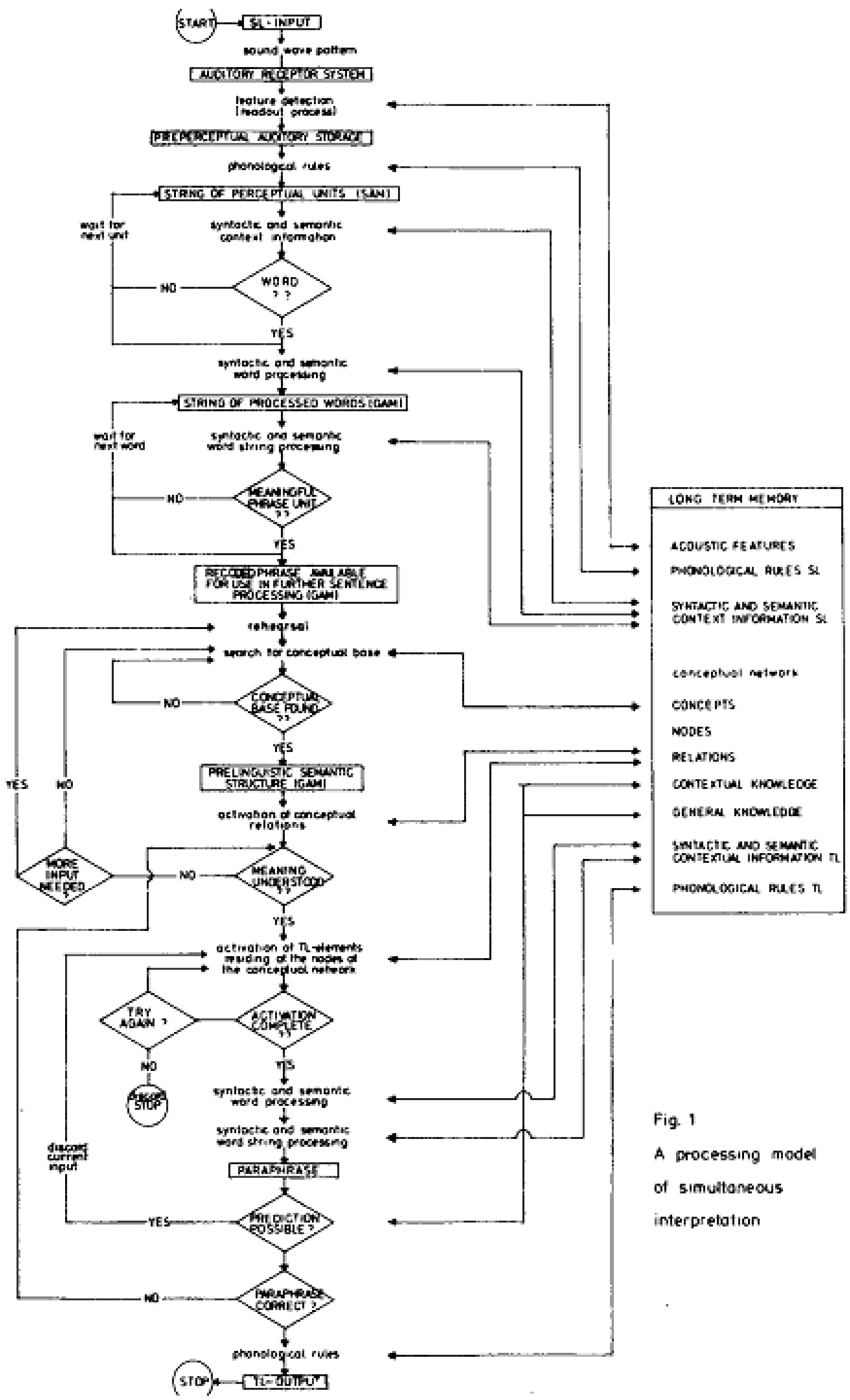

Figure 1.5 Model of SI by Barbara Moser (1978)

The third and more recent model we will briefly discuss was proposed by Valeria Darò and Franco Fabbro (Darò and Fabbro, 1994). The authors merged contemporary findings and thoughts from psychology of memory, and used them to inform their model of SI (see Figure 1.6), which – like Gerver's model – centers on memory and disregards other processes and structural entities. Among the most interesting features of the model is the fact that it is very much in line with current thinking about memory systems. The model assumes two memory systems: working memory and long-term memory, both of which are further fractionated into sub-systems. The working memory system is based on the model by Baddeley and Hitch (1974, Baddeley, 1990), but Darò and Fabbro adopt only the verbal slave system and the central executive component.

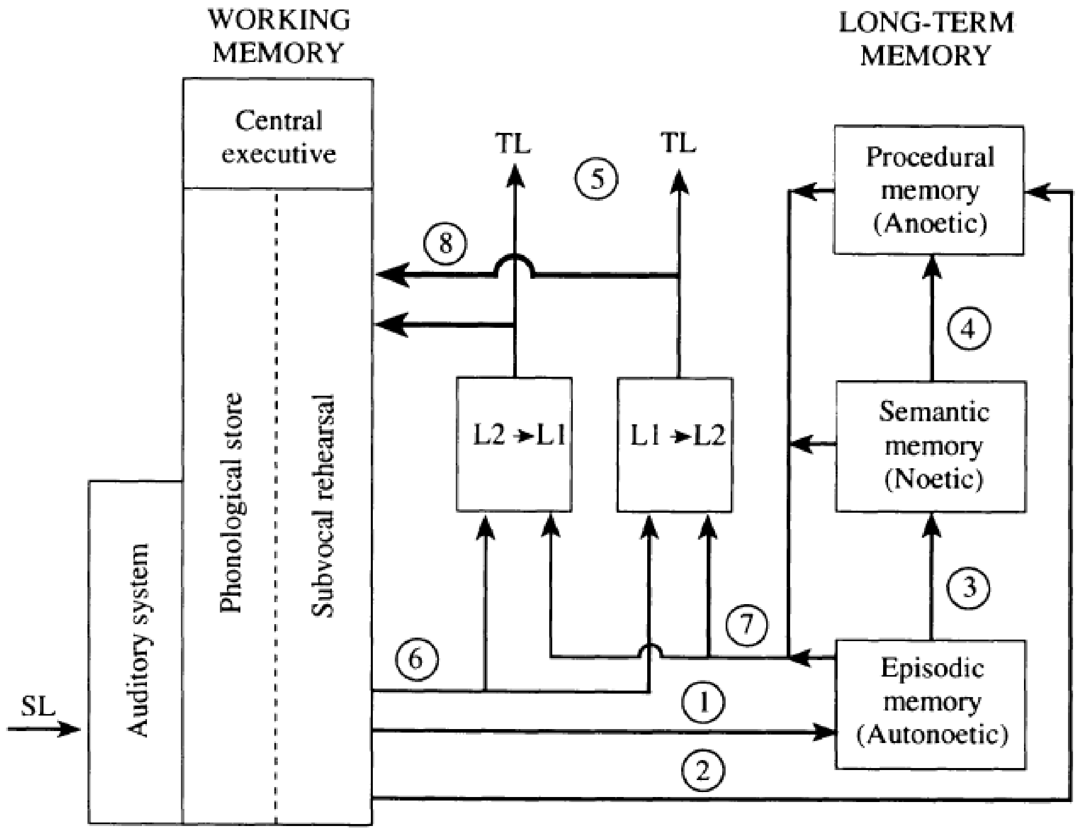

Figure 1.6 Model of simultaneous interpreting by Darò and Fabbro (1994)

Secondly, as is apparent from the model, working memory serves as a gateway for source language input. Working memory serves primarily as a passive store of the source text, but the target language interferes with this function. The target language thus limits the capacity of working memory. Thirdly, it is interesting that the authors adopted the central executive, but did not assign it any specific task or function. That may have several reasons. One, this neglect is fully in line with the state of the art in psychology at the time, when researchers were also primarily interested in the storage functions. The central executive component was proposed in the original model in 1974 (Baddeley and Hitch, 1974), but attracted more attention only in mid 1990s, after Alan Baddeley called for more research into the executive functions of working memory (Baddeley, 1996). Another reason may be that the authors were explicitly interested in aspects of verbal memory. A third interesting aspect of the model is that the actual translation processes are performed by two separate modules. Each module serves for translation in one direction – from the non-native language into the native language, and from the native language to the non-native language. It would be interesting to know whether two more modules would be needed for each non-native language an interpreter works from/to.

1.5. Working memory and its role in interpreting process models

A common feature of the three models briefly described in the previous section is that they put working memory into the centre of the interpreting process. This is, after all, true of the majority of SI process models (cf. Gile, 1995, 1997; Mizuno, 2005, see also Seleskovitch, 1968/1978). However, the models do differ significantly in their conceptualisation of the specific tasks working memory performs during the interpreting process.

The first difference is due to the period of origin of the model. All models are more or less based on the contemporary state of the art in cognitive psychology. Thus Gerver's model, which comes first chronologically, conceptualises working memory as a passive store of verbal input. Also Darò and Fabbro consider only the storage function, although their model already includes a central executive component. On the other hand, Moser (Moser, 1978), whose model is only a few years younger than Gerver's, assigns a fairly broad range of co-ordination and executive functions to working memory. It is not fully clear from the model, whether these tasks are performed by working memory (*generated abstract memory* in Moser's terminology) itself, or whether working memory serves as a kind of workspace, where the verbal material to be processed is gathered, while the actual processing is done by independent structural entities. If the latter, Moser's proposal would be similar to Darò and Fabbro's model, which assumes independent and self-contained modules responsible for the actual translation. Moser's model is thus ambiguous, but the fact that she placed these functions within working memory, is in itself a very modern concept, fully valid 30 years after the model had been proposed.

Another important difference between the models is the amount of available empirical support. As far as we know, Gerver's and Moser's models were never independently empirically tested. They represent theoretical models, based on contemporary thinking in cognitive psychology modified to reflect personal experience or partial empirical studies, but they were not submitted to a subsequent test. On the other hand, Darò and Fabbro's model is based on general theories of memory. Their specifications for purposes of SI are based on a fairly specific empirical study. Admittedly, only one study was carried out, yet among the SI models even this modest test is unique.

A third difference between the models is the proportion of the process (or amount of processing) ascribed to working memory. All models consider working memory to play an important role, but assign a differing amount of "work" to it. Gerver's and Darò and Fabbro's models assume that the interpreting process relies on working memory, long-term memory and unspecified translation mechanisms. On the other hand, Moser's model makes do with working memory and long-term memory only. Moser's model is then the only model where working memory seems to be an important mechanism directly involved in interpreting.

These differences directly affect their use in interpreting research. The most flexible, in our opinion, is Moser's model, which allows for direct testing.[4] This is not applicable to the other two models where working memory plays a central, though only a supportive role. By proposing independent translation modules, the models raise the question whether working memory is employed in interpreting above and beyond its use in normal life. This is an important point for empirical testing of the role of working memory in interpreting, which will be discussed in Chapter 3. As will be shown, most research is motivated by this very rationale: interpreters use working memory to maintain verbal material which awaits processing.

[4] Such testing would not be methodologically trivial, but the model specifies relationships between working memory and interpreting.

Chapter 2. Measuring working memory

The influential model of working memory proposed by Baddeley and Hitch (1974) offered a solution to the problem of interaction between storage and processing and replaced the older concept of short-term memory. While short-term memory serves for retention of information over short periods of time, working memory is responsible for maintenance of information *during processing*. Crucially for empirical testing of this functional definition, working memory was soon operationalised by the first complex span task, the reading span task developed by Daneman and Carpenter (1980). Previous tasks were based on strings of pieces of information, such as words or digits, which the subjects had to remember and later recall. Daneman and Carpenter's (1980) reading span task combined storage and processing by presenting participants with sentences to read and evaluate (processing task) while at the same time asking them to remember and later recall the last word of each sentence. Importantly, Daneman and Carpenter also showed that performance on the reading span task distinguishes between good and poor readers. This finding provided a very important piece of evidence, supported the usefulness of the concept of working memory and showed a fruitful way of the concept application in empirical research.

It is important to note at this point that there is a difference between working memory as a *cognitive construct* and working memory *tasks*. Working memory as a cognitive construct is a conceptual entity. Working memory is not directly observable or measurable and empirical research into working memory is crucially dependent on operationalisation of the construct by means of working memory *tasks*. As a consequence, working memory is, to an extent, what we measure by working memory tasks. The relationship between the latent cognitive construct and the tasks can be somewhat arbitrary. (This is a common issue to all entities which are not directly observable.) Since the reading span task developed by Daneman and Carpenter in 1980, a number of other working memory tasks have been developed. An important task is the operation span task developed by Turner and Engle (1989). This task uses words as the information to be remembered, and mathematical operations as the processing part of the task. The operation span task effectively uses two separate sets of materials for the storage and processing parts of the tasks. The task has proven to be robust and yields results very similar to the reading span task, suggesting that the processing versus storage parts of the task are independent of each other.

Reading span and other span tasks have been used extensively over the past 30 years and provided a large body of evidence that working memory is related to a wide range of human activities, skills and social interactions. Using this approach, working memory has been related to a large number of higher-order cognitive abilities (Feldmann Barrett et al., 2004, Engle and Kane, 2004), such as language comprehension, reading comprehension, reasoning etc. Working memory has also been shown to be crucial for cognitive development in children and linked to depression and ability to deal with stress (see Conway et al. 2005 for more examples).

However, at the same time, the exact reason for these very strong findings has not been determined (Cowan, 2000/1; Jarrold and Towse, 2006). Why exactly do working memory tasks show a relationship with other types of human performance? In other words, the findings are *descriptive*, but so far lack an *explanation*.

The research inspired by the Baddeley and Hitch's 1974 model focused very strongly on the dichotomy between verbal and spatial memory, in other words on the *memory* component of working memory. While the original model included a central executive component, there has been very little research into it until the 1990s. This is also very clearly evidenced by the structure of working memory tasks, such as the reading span and the operation span. Although the tasks are supposed to measure memory and processing, the mechanics of the task scoring are such that the outcome is based solely on the memory component of the task. For example, in the reading span tasks, participants are asked 1) to read and evaluate sentences and 2) to remember the last word of each sentence. But the score on the task is calculated as the number of words correctly recalled. The processing task – sentence reading and evaluation – is not assessed in any way.

Several studies included both the processing and the storage components in the task score. Duff and Logie (2001) report an experimental study, the results of which suggest that storage and processing are two independent components. Similar results have been reported by Bayliss, Jarrold, Gunn and Baddeley (2003). Waters and Caplan (1996) used a sentence span task, which closely resembles the reading span task, but measured the processing task as well as recall of the last words. The method of scoring both recall and processing produced better correlation with comprehension than simple reading span. Importantly, the processing task measure alone was the best predictor for reading comprehension, with small independent contribution from the recall part.

This is suggestive of some executive component being involved, rather than just the passive memory stores. Indeed, Gathercole and Baddeley (1993) concluded that most adult language processing does not require the phonological loop, which is part of the Baddeley and

Hitch verbal storage component. Instead, the authors claim, adult language processing depends more critically on the operations of the central executive.

For a long time, the central executive remained a rather underexplored component of the working memory model. Baddeley (1996) made a call for more research into the central executive, which he considers to be the most important component of working memory. Baddeley (1996) also outlined four main functions of the central executive: coordination of two tasks, switching of retrieval strategies, selective attention and stimulus inhibition, and holding and manipulation of information in long-term memory. Two important aspects of this call need to be noted. First, the central executive is not to be considered a unitary, single-function element. On the contrary, it can apparently be broken down into a number of other functions and/or components. Secondly, a typical central executive task, according to Baddeley (1996), is random number generation. In this task, participants are asked to generate a random sequence of numbers, naming them one by one. Importantly, the task does not include any explicit store-and-recall element. This indicates that central executive tasks can tap the *working* part of working memory, without necessarily tapping the memory part at all.

2.1. Components of working memory

In the past 15 years, there has been a huge increase in the number of studies focusing on exploring the central executive and the exact interplay between the storage and processing (*memory* and *working*) components of working memory. Researchers proposed a number of other executive functions. For example, Oberauer, Süss, Wilhelm and Wittman (2003) proposed coordination and supervision as the main functions, with other subfunctions distinguishable within these broad categories. Friedman and Miyake (2004) focused on the exploration of inhibition functions, whereby executive processes suppress interfering or distracting information, and distinguished between two types of inhibition: inhibition of prepotent (strongly favoured) or automatic responses and resistance to distractors on the one hand, and inhibition of proactive interference (i.e. interference from previous stimuli) on the other.

Another strand of research focuses on inter-relating the various functions and components of working memory. Similarly to the finding by Duff and Logie (2001), a number of studies point in the direction of the executive functions being separate and clearly distinguishable from maintenance (or storage) functions. Pure storage tasks typically show a weaker relationship with complex cognitive tasks (Jarrold and Towse, 2006). More complex research into the hierarchy of the tasks also suggests that tasks combining storage and processing (such as the reading span task) correlate much better than simple storage tasks with general fluid cognition (general abstract intelligence; Conway et al., 2002). On the other hand, it has been shown that individuals with large working memory capacity (high scorers on tasks such as the reading span task) do not perform significantly better than individuals with small working memory capacity in conditions where they are prevented from using the central executive functions (Engle, 2002; Hester and Garavan, 2005). This suggests that they employ the central executive to help them maintain information for later recall and that in this way the central executive contributes to their high working memory span (Engle, 2002). On the basis of these research findings, there seems to be a growing consensus that the crucial element of working memory tasks (complex span tasks) is controlled attention (also called executive attention and central executive; Engle 2002; Engle and Kane, 2004, Conway et al., 2005, Feldman Barret et al., 2004). As Engle (2002) put it, working memory capacity is not directly about memory – it is about using attention to maintain or suppress information. Working memory capacity is about memory only indirectly.

This re-evaluation of working memory properties has led to an adjustment of working memory definition. While maintenance of information in the face of other processing demands is still valid (e.g. Cowan, 2000; Caplan and Waters, 1999), it has been stressed that maintenance does not necessarily imply verbatim retention of information such as words which will be recalled later, but that it can refer to such broad concepts as maintenance of task goals (Engle, 2002). Secondly, considering the crucial role played by the executive functions, large working memory capacity does not imply a large store, but rather a better ability to keep the information active and accessible (Engle, 2002). This view means a very considerable shift away from the original focus on the *memory* component of working memory, and it will become a crucial consideration for the design of the study to be reported in this dissertation.

Engle, Tuholski, Laughlin and Conway (1999) defined the relationship between working memory and short-term memory (the traditional memory for limited period of storage and maintenance) in the following way:

WM = STM + central executive (controlled attention) + error

Their analysis of working memory and short-term memory shows that the two types of memory are highly related, yet separate constructs. Working memory, but not short-term memory, is related to general fluid intelligence *g*. Short-term memory, on the other hand, contributed to SAT[5] scores above and beyond the contribution made by working memory. Engle et al. (1999) concluded that working memory reflects more general resources, while short-term memory reflects domain-specific resources.

The notion of executive functions – and hence working memory – being domain-general now seems to receive more support (Engle, Tuholski, Laughlin and Conway, 1999; Jarrold and Towse, 2006). This has been demonstrated through the use of a variety of stimuli and processing tasks, such as reading tasks (a highly practiced activity) or arithmetic operations, all of which produce a similar pattern of results. This is taken as evidence of the general nature of the central executive and a reason for its demonstrated strong relationship to the general fluid intelligence (Cowan, 2000).

This general nature of working memory led to its classification as one of the *cognitive primitives* (Hambrick and Oswald, 2005): a basic cognitive component underlying a large number of cognitive processes. Hunt (1978) distinguished between two aspects of cognitive processing: information and the ability to manipulate information. Working memory is then considered as one of the basic abilities to manipulate information. Hunt further distinguished between processes on the basis of how much cognitive effort they require. Some processes are fully automatic and do not make any demands on attention. Among these are basic perception (visual object recognition, decoding of acoustic signal etc.) and also some highly practiced tasks, such as reading (it requires effort to look at a string of letters and *not* read them). On the other end of the spectrum are tasks which cannot be automated at all and always require effortful processing and attention. These include mostly higher-order complex behaviour, such as problem solving and reasoning. A close relationship between working memory and attention is central to Cowan's model of working memory (Cowan, 1995, 2000/1), and it implies that working memory is most important in those tasks where processing is effortful and attention demanding.

Having laid ground for closer examination of working memory, let us now turn to the relationship between working memory and some complex cognitive behaviour. Specifically, we will briefly discuss three areas, which also have a bearing on simultaneous interpretation: language processing in general, bilingualism and skilled behaviour.

[5] Scholastic aptitude test; a standardised test for high school leavers used in the USA.

2.2. Working memory and language processing

Baddeley and Hitch's 1974 model of working memory postulated a separate store for verbal information. That raises natural questions about the involvement of working memory in language processing. The relationship between working memory and various language tasks has been the subject of very intensive and fruitful research, although not all types of language processes were researched to the same extent. This strand of research was largely moulded by the general research agenda in psycholinguistics and also by the availability of methods. Most research in language processing in general has been conducted in the area of semantics at the lexical level (comparison of long and short words, concrete vs abstract words, high- and low-imagery words, homophones and homographs), sentence processing (syntactic parsing, syntactic complexity and ambiguity, integration of incoming information) and some text processes (pronominal reference, integration with pragmatic information, inferences), etc. Disproportionately more research has been carried out on comprehension than production, and on visual processing (reading) than on auditory processing (listening).

With the advent of complex span tasks, working memory was related to high-level language processing. Daneman and Carpenter (1980) have demonstrated that the reading span task, but not digit span or word span, correlates with reading comprehension as measured by the verbal SAT score, a task involving fact retrieval and pronominal anaphoric reference (but see Engle, Tuholski, Laughlin and Conway, 1999). Similarly, working memory predicts lexical ambiguity resolution (Daneman and Carpenter, 1983), paragraph comprehension (Waters and Caplan, 2005), individual ability to make use of context to derive meaning of unknown words both in comprehension and in production (Daneman and Green, 1986) or interaction of syntactic and pragmatic information (Just and Carpenter, 1992). These relationships hold when working memory is measured by complex span tasks, such as the reading span, but not when it is measured by the more traditional tasks such as the digit or word span tasks (Daneman and Carpenter, 1980). This raises the question as to which components of working memory are involved in language processing and how.

Traditional tasks such as the digit and word span are agreed to reflect storage in short-term memory rather than working memory, as there is no processing involved. Yet these tasks also proved very useful in establishing that the passive verbal store (the phonological loop in the Baddeley and Hitch's framework) is involved in vocabulary acquisition. Specifically, people with large store capacity tend to have a larger vocabulary and learn new words more quickly (Baddeley, Gathercole and Papagno, 1998).

The phonological loop seems to be at least partially involved in more complex language processes as well, although its role is secondary (Carpenter, Miyake and Just, 1995). During comprehension, a trace of the verbatim form remains in the store for some time, usually until the current unit, typically a clause or sentence, is completed (Jarvella, 1971).

The role of the passive store has been reassessed in terms of storage of strings of words until sufficient amount of information is available for semantic processing. This is no longer thought to be the case. Most language comprehension is considered to be a very highly skilled and practiced task, which for large parts runs automatically. Meanings of incoming text are automatically activated in long-term memory and if there is need for storage, it is the products of the initial processing, rather than the raw incoming text, that is entered into working memory for maintenance (Cowan, 1988; Logie, 2006).

Research into the role of phonological loop in language processing often uses articulatory suppression. According to Baddeley (2002), maintenance of information in the phonological loop is possible through rehearsal of the information. If this rehearsal is prevented, maintenance will be affected. Articulatory suppression is one effective way of such disruption. As participants are trying to remember something, for example a word list, they are asked to repeat some irrelevant material, for example say *bla bla bla*. The concurrent sub-vocalisation disrupts recall of the word list. Effects of articulatory suppression will be discussed in Chapter 3 in the context of simultaneous interpreting, a task with constant disruption by one's own voice.

The visuospatial sketchpad, a passive store for visual and spatial information, seems to play some, albeit limited, role in language processing as well. Specifically, it is used in the processing of statements with a visual and spatial element, such as *evaluating the truthfulness* (as opposed to comprehension) of sentences such as *An arm is longer than a finger* (Engle and Conway, 1998).

Complex span tasks, on the other hand, reflect both the passive stores and the central executive. Simple span tasks correlate less, if at all, than complex span tasks with higher-level language processing. This indicates that the relationship may be mostly due to the central executive component of working memory. Indeed, as mentioned above, Gathercole and Baddeley (1993) claim that most of adult human language processing does not employ the phonological loop, i.e. the lower-level storage components of working memory. This view is supported by Waters and Caplan (1996). They found that if the reading span task is scored on both recall (traditional method developed by Daneman and Caprenter, 1980) and on the processing component (which is normally disregarded), it is the processing component which provides the best prediction of reading comprehension. Similarly, Baddeley (2002) found that sentence recall is disrupted by a concurrent central executive task, while word recall is affected by disruption to the phonological loop.

But as discussed on p. 33, the central executive itself is a collective name for a host of individual functions, and it is not obvious which one or which ones should be responsible for language processing. Miyake, Just and Carpenter (1994) proposed that better performance is due to better ability to maintain information active. Gernsbacher and Foertsch (1999) on the other hand argue that better working memory means better ability to suppress irrelevant information. Engle and Conway (1998) offer an explanation which accommodates both views by proposing that the relationship is due to controlled attention – which is capable of both maintaining relevant information active and of suppressing irrelevant information. This explanation is neat and parsimonious, but goes back to the central executive being responsible without differentiating in what way it is involved.

In any case, there seems to be agreement that a substantial part of language processing is automatic (certainly in practiced users), and working memory is employed only where the processing requires some effort. Engle and Conway (1998) concluded that comprehension can proceed without employing working memory whenever the content is produced in short, simple, affirmative sentences; when there are no intratextual references (such as pronominal anaphora); when words and phrases are completely unambiguous; when the content has a linear structure; and when there are no environmental distractions. For example, Waters and Caplan (2005) found that working memory is not involved in syntactic parsing of a sentence, but is involved in semantic integration of information in paragraphs. However, when the syntax becomes ambiguous, individuals with lower working memory spans are less successful in the ambiguity resolution than persons with high spans (Swets, Desmet, Hambrick and Ferreira, 2007).

Similarly, working memory involvement in language processing will be one of coordinator rather than a primary processor. As Was and Woltz (2007) point out, the amount of information needed for tasks as complex as language comprehension far exceeds the empirically verified capacity of working memory. This clearly indicates that much of what goes on during language processing must be activated in long-term memory. Was and Woltz refer to the available information as activated long-term memory. The same proposal was made by Ericsson and Delaney (1998), who claim that reading differences reflect differences in knowledge and acquired memory skills, and that reading span is a measure of long-term memory. Also, skilled language use (as opposed to e.g. vocabulary learning) probably employs working memory for maintenance of intermediate products rather than raw language input (Cowan, 1988 Logie, 2006). Correspondingly, maintenance in working memory does not mean actual storage, but rather maintenance in active state of information in long-term memory (Engle and Conway, 1998). For example, if a sentence contains an ambiguous word, it is important to maintain all meanings of the word active until further context resolves which of the meanings is appropriate. Better performance on comprehension tasks is then achieved by individuals who are better able to maintain more information active, and who can effectively select and/or suppress what is relevant and irrelevant.

2.3. Working memory and bilingual language processing

A specific type of language processing is involved in bilingual language use. Working memory is not among the most important issues in bilingual research, nevertheless there are selected topics which have a bearing upon the topic of this dissertation. These include second language learning and language control.

It has been already discussed on p. 36, there is strong evidence of the involvement of the phonological loop in vocabulary acquisition (Baddeley, Gathercole and Papagno, 1998). This is especially relevant for second-language acquisition in late bilinguals (people who learned the second language post-puberty). Other evidence suggests that the relationship between working memory and language processing in learners of a foreign language is similar to that between working memory and language processing in one’s native language. Specifically, working memory seems to be employed especially in more complex tasks; it can further serve as a facilitator – learners with larger working memory capacity seem to learn the second language faster (Miyake and Friedman, 1998).

The second issue is language control. Several models have been proposed to explain how bilinguals control access to the languages. The first question is whether languages in bilingual persons are activated selectively or whether both of the languages are activated at the same time. Abundant empirical evidence from experimental work shows that bilinguals activate both languages in parallel both during comprehension and production (Colomé, 2001; Kroll, Bobb, Misra and Guo, 2008). The issue of production then becomes how bilinguals control in which language they will produce their output. Two important competing models have been proposed. The inhibitory control model (Green, 1998) assumes that both languages are activated, candidates (e.g. words) from either language compete for production, and selection of the language to be produced is then achieved through inhibition of the unintended language. The second model proposes that bilinguals acquire (and improve) the ability of selective attention, which allows them to focus on one language only. A recent review by Kroll et al. (2008) seems to side for the inhibitory control hypothesis, but the authors also concede that there is a large number of unresolved issues related to context of use, age of acquisition, balance between the two languages, their similarity etc. It also needs to be pointed out that for methodological reasons, most of the research carried out in the area of language control uses the method of single words. The exact nature of language control is not the topic of this dissertation, but we note with interest that either of the two models makes use of some form of attentional control. Indeed, research by Bialystok and Craik (2010) suggests that bilinguals may have a more advanced executive control system compared to monolinguals. According to Bialystok and Craik, this allows them to be less sensitive to distractions, be able to select relevant aspects of the information presented to them etc. Additionally, these advantages carry over to non-linguistic tasks. On the downside, bilinguals seem to have a smaller vocabulary in either language than monolinguals and reduced (slower) access to lexical items (Bialystok and Craik, 2010).

Research on bilingualism seems to suggest that there is an advantage of bilingualism over monolingualism in better functioning of the executive control, although it is not clear yet what the exact mechanisms behind this “bilingual advantage” are (Bialystok and Craik, 2010).

2.4. Working memory and skilled performance

One of the models of working memory discussed in Chapter 1 was specifically developed to address the question of skilled and expert performance. The long-term working memory model by Ericsson and Kintsch (1995) is a response to empirical findings in which short-term memory content highly exceeds the established limit on memory store size. The model also explains differences in performance among individuals with varying amount of experience in a given domain. Crucially, long-term working memory is not a general construct, but is specific to an area of highly practiced activity.

The long-term working memory model allows highly skilled performers to retain information (or keep it active) for long periods of time, in amounts in excess of the traditional capacity, and also in a robust state whereby the information is fairly unaffected by distractors, such as an intervening task (Ericsson and Kintsch, 1995). Findings in support of the model come from domains as diverse as comparison of novice and expert chess players on the one hand, and from tasks as routine as reading on the other (Ericsson and Kintsch, 1995).

On the other hand, involvement of working memory in skilled behaviour seems to decrease with increasing skill. Skilled behaviour is assumed to be acquired in three stages: the cognitive stage, the associative stage, and the autonomous stage (Matthews, Davies, Westerman and Stammers, 2000). The acquisition process starts with conscious processing of each step of the task, proceeds through better control and coordination until the autonomous stage is reached where the task can be performed automatically. Ackerman (1988) investigated the three stages in several tasks, including an air traffic control simulation task. Figure 2.1 shows his findings.

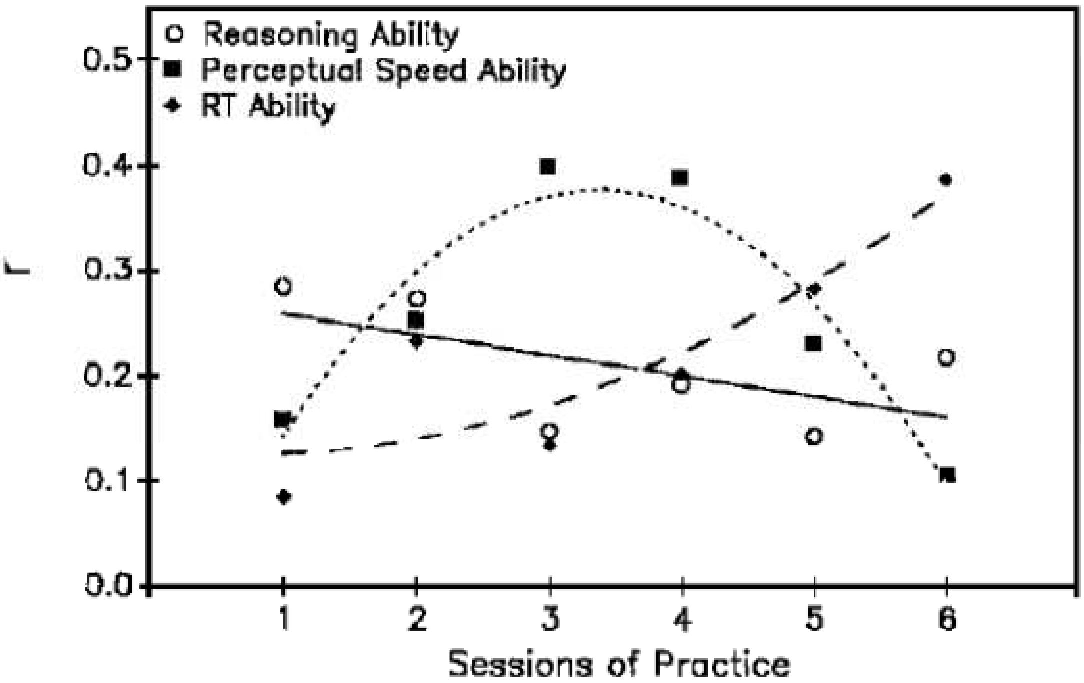

Figure 2.1 Predictors of performance with increased practice (Ackerman, 1988)

In the initial stage of the skill acquisition, reasoning ability was the most important predictor. In the second stage, perceptual speed was the strongest predictor, and in the final stage, performance was mostly determined by psychomotor speed. These findings show that the involvement of control mechanisms, such as the central executive, are strongest in the initial stages of task acquisition. At that point, the individual has no routines to fall back on and each step towards the task completion requires reasoning, decision making, problem solving, in other words – it demands attention and control. However, as routines are developed the involvement of such executive mechanism becomes less important. It is important to note that Ackerman's findings would hold for tasks where automation is possible. As long as the task can be completed relying on developed routines, executive control may not be necessary. However, tasks which cannot be fully automated will always require attention and control.

While Ackerman focused on changing performance predictors in skill acquisition, Woltz (1988) addressed specifically the question of working memory. He found a relationship between working memory and skilled performance both in the initial stages of skill acquisition and in later stages when task performance approached automation. However, the nature of working memory involvement was different. Like Ackerman, Woltz found that in the initial stages attentional control was key to task performance. In later stages, the role of attentional control decreased, but the need for information activation increased. As a task becomes complex, successful performance requires more information to be simultaneously maintained in an active state. This finding broadly agrees with the long-term working memory model, which explains how skilled and expert performers maintain complex patterns of information active. The role of general working memory and long-term working memory was explicitly tested by Sohn and Doane (2003). Their findings support the involvement of working memory in skilled task performance along the lines described above: general working memory is important in the performance of less skilled individuals. Those scoring high on the long-term working memory task (i.e. those with more experience and capable of more complex task representation) relied more on their long-term working memory. Additionally, there is evidence that working memory is a contributing factor to skilled behaviour above and beyond task specific skills. Hambrick and Oswald (2005) found domain-specific knowledge about baseball to be the main predictor of performance in recall of baseball games, but working memory also made an independent contribution. Similarly, Meinz and Hambrick (2010) have shown that both extensive deliberate practice (Ericsson, 2000) and working memory explain independently variation in expert performance in music sight reading.

These findings provide support for the employment of both general working memory and task-specific long-term working memory at different levels of acquired skill. The initial stages are more sensitive to the ability to control attention, while in later stages, the ability to activate task-relevant information and to maintain the information active is crucial. It needs to be noted here that testing long-term working memory requires understanding of the underlying processes of the given skill, and also the development of skill-specific LTWM tasks.

In this chapter, we have discussed a more detailed specification of working memory. Working memory is a multi-component construct and different components are employed to varying extent in different higher-level cognitive tasks. As a latent construct, it is empirically accessible only through working memory tasks. Quintessential working memory tasks, such as the reading span, show excellent validity and reliability in terms of their relationship to a range of human activities, but the exact nature of the relationship is not known. Understanding the involvement of working memory in complex skilled behaviour, such as interpreting, requires a more detailed identification of specific working memory components/functions, and selection of appropriate tasks to tap these components. Against this general background, we will now discuss previous research on the involvement of working memory in simultaneous interpreting.

Chapter 3. Simultaneous interpreting and working memory: previous research

Working memory is probably the single most often researched isolated cognitive component in interpreting studies. The first empirical studies dedicated to interpreters' working memory date back to the 1990s (Darò and Fabbro, 1994, Padilla, Bajo, Cañas and Padilla, 1995), although an interest in interpreters' working memory goes back to early interpreting research and theories. Seleskovitch (1968/1978) considered excellent memory to be the cornerstone of interpreting, and Gerver (1975) built his interpreting process model around short-term memory stores. Good working memory came to be considered the very basis of the interpreting skill (Darò, 1989; Bajo, Padilla and Padilla, 2000). Empirical research has therefore mostly focused on comparing working memory between interpreters and non-interpreters (interpreting students, non-interpreter bilinguals) in search of evidence that interpreters' working memory capacity is larger than that of non-interpreters (Padilla et al., 1995, Christoffels, de Groot and Kroll, 2006). A subgroup of studies focused on a specific mechanism of working memory, the phonological loop, and the way storage is disrupted through concurrent articulation. The assumption behind all these studies is that if interpreters can be shown to have better working memory than individuals with similar background (education, bilingualism, age), then better memory could be attributed to interpreting. If this advantage held for comparison of professional interpreters to interpreting students, then better working memory would be better the result of extensive practice rather than aptitude. This is the basic rationale behind the majority of working memory studies conducted to date.

A second strand of research involves empirical investigations in the context of the interpreting task itself. There are only a few studies, as far as we know, and each takes a very different approach. Shlesinger (2000) tested specific predictions, derived from working memory research, in the context of simultaneous interpreting, Liu (2001) compared interpreting students' and professionals' performance on a working memory task and an interpreting task. Christoffels et al. (2003) related working memory to simultaneous interpreting in untrained bilinguals, Tzou et al. (2011) compared working memory and simultaneous interpreting in untrained bilinguals and interpreting students. Hodáková (2009) tested working memory in two groups of interpreting students with a different level of training.

To our knowledge, a total of thirteen empirical studies are available with working memory and interpreting and/or interpreters as their main focus. More than half of the studies are doctoral dissertations[6], attesting again to the importance of working memory in interpreting research. The above classification of studies into the two broad categories – focus on interpreters and focus on interpreting - will serve as a framework for a more detailed discussion of the research carried out and published to date.

3.1. Do interpreters have a better working memory than non-interpreters?

A comparison of interpreters and non-interpreters (and possibly interpreting students) in terms of working memory has been the primary or secondary focus of nine studies. The hypothesis that interpreting practice enhances working memory was not borne out conclusively. Some studies report superior working memory in interpreters, others have found no difference between groups, or found a difference in the opposite direction (Köpke and Nespoulos, 2006) with students performing better than any other group. In Chapters 1 and 2, we elaborated on the construct of working memory, showing that it comprises several components with different functions, broadly divided into storage and central executive. Correspondingly, previous research can be discussed according to the working memory function it targeted: storage, storage and processing, and central executive.

The storage function of working memory can be measured by traditional short-term memory tasks. Storage tasks previously used in interpreting research include the digit span, the word span and the non-word span tasks.[7] Table 3.1 provides a summary of studies using these tasks (and other working memory tasks to be discussed next). In the digit span task, participants are presented with ever longer strings of digits and then asked to recall them. The word span and non-word span tasks are based on the same principle, but use words or word-like strings. Word span task therefore contains an element of semantic processing, while non-words (strings of letters which conform to the phonological rules of a language, but do not have any semantic content) place more demands on phonological memory. Finally, in the free recall tasks, participants are again presented with lists of words, but there are two differences with the word span task: all lists have a fixed length, e.g. 15 words, and the words can be recalled in any order.

[6] Not all doctoral dissertations were available in their original upublished format. The following discussion typically refers to articles published on the basis of the dissertation. Some dissertations were published in more than one article.
[7] A number of other tasks were used as well (see e.g. Köpke and Nespoulous, 2006; Signorelli et al., 2011), but they typically produced no differences among the tested groups and were used in a single study only.

Table 3.1 Overview of working memory tasks used in interpreting research

Task	Study	Main findings
Storage tasks		
digit span	Padilla et al., 1995	interpreters performed better than interpreting students and controls
	Chincotta and Underwood, 1998	no difference between interpreting students and controls
	Köpke and Nespoulous, 2006	no difference between interpreters, interpreting students and controls
	Tzou et al., 2011	second year interpreting students performed better than controls
word span	Christoffels et al., 2006	interpreters performed better than controls
	Köpke and Nespoulous, 2006	no difference between interpreters, interpreting students and controls
non-word span	Köpke and Nespoulous, 2006	no difference between interpreters, interpreting students and controls
	Signorelli et al., 2011	interpreters performed better than controls
free recall	Padilla et al. 1995	no difference between interpreters, interpreting students and controls
	Köpke and Nespoulous, 2006*	no difference between interpreters, interpreting students and controls
Storage and processing tasks		
reading span	Padilla et al., 1995	interpreters performed better than interpreting students and controls
	Christoffels et al., 2006	interpreters performed better than controls
	Signorelli et al., 2011	interpreters performed better than controls
	Tzou et al., 2011	interpreting students performed better than controls
listening span	Nordet and Voegtlin, 1998	no difference between interpreters, interpreting students and controls
	Liu et al., 2004	no difference between interpreters and interpreting students
	Köpke and Nespoulous, 2006	interpreting students performed better than controls
	Timarová, 2007	interpreters and interpreting students performed better than controls
Central executive tasks		
Stroop test	Köpke and Nespoulous, 2006	no difference between interpreters, interpreting students and controls

* Köpke and Nespoulous classify free recall as a task involving storage and processing

Overall, each of the digit span, word span and non-word span tasks supported the hypothesis of superior interpreter performance on working memory tasks (‘the superiority hypothesis’) in about half of the studies and did not provide any evidence of such superiority in the other half. However, two studies (Padilla et al., 1995, Signorelli et al., 2011) report results with some qualification. Padilla et al. report a significant effect but do not provide pair-wise comparisons, so it is not clear what the results mean. Interpreters achieved a higher score than all other groups (interpreting students and non-interpreter controls), but the significant effect may reflect a difference between interpreters and students, but not necessarily between interpreters and non-interpreters. Signorelli et al. found a difference in favour of interpreters in a non-word span task after excluding outliers, but state that if different exclusion criteria are chosen, the difference is no longer significant. All three storage tasks, along with three other tasks, were used by Köpke and Nespoulous (2006), who did not find a group difference on any of them, and concluded that "short-term retention in the phonological loop does not play a major role in simultaneous interpreting" (p.11).

The last of the storage tasks listed in Table 3.1 free recall did not show any significant differences between interpreters and non-interpreters in its pure form. However, the task was included in both studies with a different purpose: studying recall of interpreters and non-interpreters under the very specific condition of articulatory suppression. Maintenance of information in memory requires the use of the phonological loop (Baddeley, 2000) and disruption of the phonological loop can severely affect memory and recall (Baddeley, 2000). More specifically, remembering over a short period of time requires rehearsal, which is subvocal in nature, and by preventing such subvocal rehearsal, recall of stored information is impaired. This effect is called articulatory suppression and in experimental research is usually achieved by requiring participants to repeat aloud a simple syllable, such as "ba ba ba". Simultaneous interpreters work under the conditions of articulatory suppression all the time, producing their own output in the target language, while listening to the input in the source language at the same time. An interesting question arises whether interpreters are more immune to the effects of articulatory suppression than other individuals, i.e. whether interpreters' recall of stored information will be affected less or not at all when asked to remember information while talking at the same time.

Padilla et al. (1995), Padilla, Bajo, and Macizo (2005) and Köpke & Nespoulous (2006) asked several groups of participants – interpreters and various groups of controls – to do the free recall task under two conditions: listening only and articulatory suppression. When interpreters and controls listened to word lists and recalled them, there was no difference between the groups. When, however, they listened to the word lists while simultaneously speaking (repeating an irrelevant syllable), there was a difference. Padilla et al. (1995) found that interpreters performed better than interpreting students and controls, and Padilla et al. (2005) found that interpreters performed better than two groups of controls. Köpke and Nespoulous (2006) found that student interpreters performed best, including better than professional interpreters. In another study, Chincotta and Underwood (1998) tested interpreting students and controls on a bilingual digit span. The digit span is sensitive to language – in different languages, the average digit span differs due to the difference in digit names length (Naveh-Benjamin and Ayres, 1986) and the authors expected that this language effect on digit span would disappear under articulatory suppression, as is usually the case. The controls behaved as expected, but the interpreting students retained a language effect, i.e. their digit span was different in the two languages indicating that the expected disruption due to articulatory suppression did not take place. All this points in the direction of interpreters somehow coping with the effects of remembering and talking at the same time. They are not immune – their performance is generally worse under articulatory suppression (listening and repeating aloud an unrelated syllable or word) than in normal listening condition but they seem to be affected less than non-interpreters.

Another series of studies compared recall in conditions of articulatory suppression and in interpreting. Darò and Fabbro (1994) found that interpreting students' digit span was worse after simultaneous interpreting than after a simple articulatory suppression, which in turn was worse than after shadowing. Gerver (1974) compared recall of text after simultaneous interpreting and shadowing in interpreting students, and found better performance in simultaneous interpreting. Christoffels (2004) found that untrained bilinguals recalled a short text after simultaneous interpreting and shadowing equally well, and better than after articulatory suppression. All these results are intriguing and will require further research. Chincotta and Underwood (1998) and Darò and Fabbro (1994) attributed the effect to divided attention, while Christoffels (2004) and Padilla et al. (2005) argued for a better interface with long-term memory (via the episodic buffer specified in Baddeley's (2000) revision of his working memory model). One point, however, should be stressed. Most of the studies tested interpreting students or graduates without substantial experience in professional simultaneous interpreting. In fact, the only study using a group of experienced interpreters was the study by Köpke and Nespoulous (2006), where no superiority of interpreters was found. The evidence for better resistance to articulatory suppression is nevertheless robust and may be usefully explored in the context of interpreter training or aptitude. In fact, Christoffels (2004) asked untrained bilinguals to simultaneously interpret a short text, and found that performance on the SI task positively correlated with recall under articulatory suppression, i.e. those who resisted the negative effects of articulatory suppression and recalled more words also produced better interpretation. So even if the exact nature of this peculiarity in working memory subfunction in interpreters is not properly understood, it may have some practical uses in applied settings.

We now move on to the second group of tasks, those combining storage and processing. The tasks in this category are considered the quintessential working memory tasks. Table 3.1 shows that two tasks were used predominantly: the reading span task, which is the original working memory task developed by Daneman and Carpenter (1980), and its variant, the listening span task. The two tasks differ very little. The reading span task consists of visually presented sentences the subjects are asked to read aloud. In the listening span task, the presentation is auditory and subjects are typically asked to indicate meaningfulness of the sentence. In both tasks, participants need to remember the last word of each sentence and then recall all the words in a given set at the end. The equivalence of the two tasks has been demonstrated in a general context (Daneman and Carpenter, 1980), and also within interpreting studies research (Köpke and Signorelli, 2011). As is apparent from Table 3.1, tasks combining storage and processing provide, in general terms, more robust evidence for interpreters' superior working memory. Four studies using the reading span (Padilla et al., 1995, Christoffels, de Groot and Kroll, 2006, Signorelli et al., 2011, Tzou et al., 2011) all found that interpreters perform better than controls. In the four studies using listening span as a measure, two did not find any advantage of interpreters (Nordet and Voegtlin, 1998, Liu et al., 2004). Köpke and Nespoulous (2006) found that interpreting students, but not professional interpreters, performed better than controls. Timarová (2007) found that both interpreters and students performed better than controls. This apparent difference between the reading and listening span tasks is puzzling. As stated above, they are considered to be very similar measures. Köpke and Signorelli (2011) analysed the discrepancy from a methodological point of view, comparing such parameters as participants, modality of presentation and scoring methods. [8] One aspect we would like to emphasise here is the scoring method.

The traditional scoring method for span tasks consists of taking the length of the longest correctly recalled string. If the subject correctly recalls all final words in a set of four sentences, but not in a set of five sentences, the span will be four. However, this method produces a very restricted range of scores: most participants will score somewhere between 3 and 6. Conway et al. (2005) have demonstrated that this method is not very sensitive, and suggest other ways of determining the score. One of the ways is to count the total number of words correctly recalled over all the trials. All of the studies in interpreting used either the method of the traditional span or the method of total words. Among the reading span studies, only Padilla et al. (1995) used the span method, while the remaining three studies used the total words method. In listening span studies, Nordet and Voegtlin (1998) used the method of

[8] Their thorough and interesting analysis concerns discrepancies in other working memory tasks as well.

total words, Köpke and Nespoulous (2006) and Liu et al. (2004) used the span method. Timarová (2007) calculated span using both the traditional span method and the total words method. The traditional span method did not show any difference between the groups, the total words method demonstrated better performance of interpreters and interpreting students. While this is only one possible reason (see Köpke and Signorelli, 2011, for a much more extensive discussion), it is a plausible explanation for the apparent difference between reading and listening span studies in interpreting.

In any case, the storage and processing tasks seem to provide much less ambiguous evidence for the advantage of interpreters than the storage tasks. Since the difference between them is assumed to be in the involvement of the central executive component of working memory, let us now turn to research involving central executive tasks. Table 3.1 lists only one such study. Köpke and Nespoulous (2006) tested interpreters, interpreting students and controls on the Stroop test (Stroop, 1935). This classic test works in the following way: participants are presented with a series of words, and their task is to simply name the colour in which the words are printed. Figure 3.1 shows three examples. In a), participants see a neutral word, such as "house" and name the colour in which it is printed, black in this case. Apart from such words, the test includes a number of words which themselves are colour names. These are printed either in matching ink, as in b), where the word "black" is printed in black ink, or in ink of a different colour, as in c), where the word "white" is printed in black. Since reading is a highly overpracticed activity, people have problems *not* reading the word, and in conditions depicted under c) will often make mistakes by reading the word printed, rather than naming its colour.

Figure 3.1 Stroop test: neutral word (a), colour word in matching (b) and non-matching ink (c)

The Stroop test is considered to measure inhibition of this automatic, or prepotent, response (Miyake et al., 2000). Köpke and Nespoulous (2006) administered the test in four versions – monolingual English, monolingual French, and bilingual English-French (both directions). Of the four versions, the only difference found was a superior performance of interpreting students on the French bilingual version (English colour words to be named in

French), in comparison to both professional interpreters and bilingual controls. However, the authors do not attribute the difference to better functioning of the central executive, but rather to less balanced language proficiency in the group, which would make them less vulnerable to interference from the English written words.

To summarise, in this section we reviewed previous empirical research in interpreting studies which addressed the question of whether interpreters have better working memory than non-interpreters. The general approach has been to compare interpreters to control subjects on a variety of working memory tasks. The overall broad conclusion is that interpreters do not seem to outperform other individuals, including interpreting students, on simple storage tasks (short-term memory tasks) where memorising and recall are required. While some studies did report evidence in favour of interpreters' better performance, these findings are sometimes qualified by methodological or reporting parameters. Köpke and colleagues concluded that "maintenance rehearsal […] probably plays only a minor role in expert interpreting" (Köpke and Signorelli, 2011:13, also Köpke and Nespoulous, 2006). Tasks combining storage and processing, on the other hand, provided more support for the hypothesis of better working memory in interpreters. The majority of studies found an advantage in interpreters, although here too methodological decisions with regard to participants, materials or procedure may play a role. Additionally, interesting evidence comes from the study of the effect of articulatory suppression on interpreters' recall. While interpreters did not differ from non-interpreters on recall under normal conditions, they tended to be less affected by simultaneously speaking than non-interpreters. These studies, however, usually did not test professional interpreters with substantial experience, and the findings are thus so far limited to untrained individuals, interpreting students and interpreters with fairly limited professional experience. Together with more robust findings for tasks combining storage and processing, these results point to the importance of the central executive processes. These, however, have not been systematically tested: only one study employed a central executive task, the Stroop test, but did not find any group differences. Finally, some of the findings on the effects of articulatory suppression were interpreted as reflecting the ability to divide attention and direct it, at least partially, away from the disruption of one's own voice, which again points in the direction of greater role of the central executive resources. One study (Christoffels, 2004) even related resistance to articulatory suppression to better performance in interpreting. In the research reviewed so far, the focus was on interpreters, with interpreting playing very little role. Let us now turn to the interpreting performance itself.

3.2. Is working memory related to interpreting?

The second research line identified at the beginning of this chapter focuses on relating working memory to interpreting. Fewer researchers pursued this question than the question of enhanced working memory, and an overview of the available studies, together with the main findings, is presented in Table 3.2.

Table 3.2 Overview of studies testing relationship between working memory and interpreting

Author (Year)	**Aim**	**Main findings**
Shlesinger, 2000	test of various predictions made by the Baddeley model	word-length effect retained at higher presentation speed recency effect – last modifier retained more often than preceding modifiers in nominal strings
Liu, 2001	effect of experience and WM on SI	professional interpreters and students equal on a measure of working memory; professionals outperformed students on an SI task
Christoffels, 2004	relationship between performance on an articulatory suppression task, working memory task and a lexical task, and an SI task	performance under articulatory suppression predicted SI performance in untrained bilinguals performance on a lexical task, reading span and digit span predicted SI performance in untrained bilinguals
Hodáková, 2009	relationship between working memory, attention and interpreting	working memory related to consecutive, but not simultaneous interpreting, in students no relationship found between attention and interpreting
Tzou et al., 2011	relationship between working memory and SI	reading span and digit span predicted SI performance in interpreting students and untrained bilinguals

In the first of these studies, Shlesinger (2000) tested a number of predictions made by the Baddeley model of working memory in the context of simultaneous interpreting. The study is quite complex and tests a large number of hypotheses, not all of which will be discussed here. Shlesinger focused primarily on the effects observed in the functioning of the phonological loop and verbal short-term memory. She asked professional interpreters to simultaneously interpret several texts, recorded at two different speeds of 120 and 140 words per minute (wpm), with embedded nominal strings consisting of a noun and four adjective modifiers, such as *constant, active, conscious, healthy interest*. The strings were of two types, consisting of bisyllabic or tetrasyllabic words. The language combination was English-Hebrew, where English is a head-final language (i.e. adjective + noun), whereas Hebrew is a head-initial language (i.e. noun + adjective). Successful interpretation therefore required that interpreters wait for the noun before starting their interpretation of the string and maintain the modifiers in memory. In line with Baddeley's model, Shlesinger predicted a word-length effect, where longer words are retained less successfully than shorter words, and recency effect, where items presented as last are recalled first. In the context of her study, this leads to two predictions. First, that tetrasyllabic strings would be recalled less successfully, i.e. with higher error rate and/or less completely than bisyllabic strings. Secondly, it was predicted that the last of the four modifiers (*healthy* in the example given) would be interpreted first and/or more frequently than the other modifiers. Additionally, it was predicted that the slower presentation rate (120wpm) would affect retention more than the fast presentation rate (140wpm) due to the need to store the information longer before interpreting it. The predictions were supported only partially. Shlesinger found that longer strings were not interpreted less completely or with more errors than short string (main effect), which suggests that the word-length effect was eliminated as is usual in conditions of concurrent articulation (articulatory suppression). However, in the fast presentation rate, tetrasyllabic strings were interpreted more successfully than bisyllabic strings. This indicates that there is a word-length effect, despite the articulatory suppression. A similar finding was reported by Chincotta and Underwood (1998, discussed in 3.1), and Shlesinger's results thus provide another piece of evidence for resistance to articulatory suppression, this time in experienced professional interpreters.

The second prediction concerned the recency effect – a preference for producing first the modifier which was presented as last. Shlesinger reports that this prediction was supported, i.e. the last modifier was retained more often than the preceding three modifiers, and often appeared as the first modifier in the interpretation. This finding supports the idea that general working memory phenomena, such as the recency effect, apply also to interpreting.

Both these results require qualification. The strings presented consisted of a noun and four modifiers. The interpretation, on the other hand, retained on average about one modifier per string. In terms of the word-length effect, this means that the amount of information stored in memory before production was still twice as long in terms of form (four versus two syllables), but identical in terms of content (one adjective). In terms of the recency effect, it is difficult to speak of producing the last presented modifier as the first interpreted modifier when, as it seems, it often was the only modifier retained. That said, Shlesinger's study is an important example of the complexities involved in applying cognitive concepts generally, and working memory specifically, to simultaneous interpreting.

The other four studies listed in Table 3.2 share a common design, where working memory and interpreting are tested separately from each other. Participants performed two types of tasks: one measuring their working memory, and the other measuring their simultaneous interpreting performance. The studies differ in the type of participants they tested, the working memory tasks used and the way they measured interpreting performance, but the broad design similarities provide sufficient ground for comparison, and the four studies will therefore be discussed together.

Very interestingly, the studies lend themselves to a useful comparison on the basis of the participant characteristics. Christoffels (2004) tested untrained bilinguals, Tzou et al. (2011) compared interpreting students and untrained bilinguals, Hodáková (2009) tested interpreting students, and Liu (2001) compared interpreting students and professional interpreters. That means that among the four studies, the full range of interpreting experience, from untrained bilinguals to professional interpreters, was submitted to a test. Table 3.3 provides an overview of the main design features and results.

Table 3.3 Relation between working memory and simultaneous interpreting at different skill levels

	Untrained	**Untrained + Students**	**Students**	**Students + Professionals**
Study	Christoffels, 2004	Tzou et al., 2011	Hodáková, 2009	Liu, 2001
Test WM	digit span	digit span	listening span	listening span
	reading span	reading span	arithmetic addition	
	articulatory suppression test		attention	
Test SI	accuracy of selected sentences + overall quality	accuracy of selected sentences + overall quality	accuracy of idea units	accuracy of idea units
Relationship between WM and SI found?	Yes	Yes	Yes	No

In two separate experiments, Christoffels (2004) tested a group of untrained bilingual students on a digit span task, reading span task and a test of articulatory suppression effect. The participants also simultaneously interpreted a short text from English into Dutch (from their second language to their mother tongue). The interpretation was scored in two ways: selected sentences were scored for their accuracy, and an overall assessment of the interpretation was scored (only the first measure was used in the experiment involving articulatory suppression). Christoffels found positive correlation for all measures. Higher resistance to articulatory suppression was associated with better performance in the interpreting task. Digit span correlated positively with both measures of interpreting (selected sentences and overall quality) and reading span correlated positively with accuracy of selected sentences. Tzou et al. (2011) tested three groups of Chinese – English bilingual participants: untrained bilinguals, interpreting students in their first year of training and interpreting students in their second year of training. They too administered a digit span and a reading span tasks (in both English and Chinese), and measured simultaneous interpreting performance in the same way as Christoffels on selected segments and as an overall evaluation. Tzou et al. report that both measures of simultaneous interpreting correlated positively with English and Chinese reading span, and English digit span.

Hodáková (2009) tested a large group of beginning and advanced interpreting students and compared their performance on a listening span task, a test of simple arithmetic addition, a test of attention, and consecutive and simultaneous interpreting (German – Slovak). She found a correlation between the listening span and consecutive interpreting and between the arithmetic addition test and simultaneous interpreting. The attention test was not related to either interpreting mode. It needs to be said that the administration of the tasks was unorthodox, as both the listening span task and the arithmetic addition test were administered at the same time – participants completed the addition problems while listening to the sentences presented as part of the listening span task. This makes the interpretation of the results more difficult. The correlation of the arithmetic test with simultaneous interpreting may possibly be due to an ability to work on two tasks at the same time. Hodáková (2009) does not offer any explanation for the decision to administer both tasks at the same time. The correlation between the two tests is positive, although statistically non-significant, which may be considered as an argument against a trade-off between the two tasks (focus more on one at the expense of the other) and an argument in favour of the combined administration testing participants' ability to divide attention. This view is supported by a respectable positive and significant correlation between the combined score (for the listening span and arithmetic task) and the attention test.

Finally, Liu (2001) tested three groups of Chinese-English interpreters: beginning and advanced interpreting students and experienced professionals. The listening span test was used as a measure of working memory, and the three groups performed very similarly on the test, with no significant differences found. On the simultaneous interpreting task, Liu administered several texts and measured accuracy of selected manipulated segments. Specifically, the segments consisted of essential and secondary idea units, followed by a continuation sentence, and embedded in three different texts. Additionally, each segment was produced in an easy and difficult version (based on a readability index), resulting in six texts. Each of the texts was recorded at two different speeds, thus making a total of 12 experimental texts administered to each participant. Liu measured two variables: correctly interpreted idea units and correctly interpreted continuation sentences as a function of importance, difficulty, speed and interpreting experience. Most importantly for our analysis, Liu found significant differences between the groups on the number of idea units correctly interpreted, with professional interpreters achieving higher score than students on the total number of units. Professional interpreters, but not students, also interpreted more essential idea units than secondary idea units, indicating that professionals are better at discriminating between the important and less important in a message and at managing better their resources under pressure by making relevant choices about what to interpret and what to leave out. Other variables yielded results in the expected direction: difficult segments were interpreted less correctly than easy segments and continuation sentences were more often correct after an easy segment than a difficult one. Interestingly, only one of the three basic texts was affected by speed. None of these findings was moderated by interpreting experience – professional interpreters were not less affected by segment difficulty or speed. Liu's conclusion was that the observed differences in simultaneous interpreting, both quantitative (more segments correctly interpreted by professionals) and qualitative (better selection of essential over secondary idea units), cannot be attributed to general cognitive ability, as the groups performed equally on the listening span test, and that interpreters must therefore draw on task-specific skills and strategies.

Taking the four studies together, a relationship between working memory and simultaneous interpreting was found in three of them – in untrained bilinguals and interpreting students, but not in professional interpreters. The finding that working memory is a predictor at lower levels of acquired skill is consistent with Ackermann (1988; discussed in Chapter 2), whereby working memory plays an important role during the process of acquisition, where automatic routines have not yet been established and where attentional control is needed to guide performance. A relationship between working memory and simultaneous interpreting in professional interpreters was tested in one study and was not demonstrated. Given the paucity of research, Liu's conclusion about general cognitive ability not being crucial might be premature.

Therefore, let us consider some alternative explanations of her results. As discussed in section 3.1, the listening span task can be scored using a variety of scoring methods. Liu used the traditional span method, which yields a very restricted range of scores (two to five points in her case) and which has a demonstrated lower sensitivity (Conway et al., 2005). There is hence a possibility that if the score were calculated using a different method, the differences between groups would be significant. Liu (2001:30), however, was not aiming to demonstrate better working memory in professional interpreters – her prediction was precisely that the groups would not differ on their general cognitive ability, as measured by the listening span, but would differ on "simultaneous interpreting, which was considered to represent their working memory performance on a domain task" (Liu, 2001:29-30). Hence Liu assumes a relationship between working memory and simultaneous interpreting by definition, although her concept of working memory is skill-related, different from the general cognitive ability, and further unspecified. This makes the evaluation of her study more difficult.

A second methodological point has to do with the fact that Liu tested general working memory and simultaneous interpreting separately. She compared the three groups on a listening span task, and concluded that the three groups did not differ. Then she tested them on a simultaneous interpreting task, and found a difference in performance. The performances on the two tests, the listening span task and the interpreting task, however, were not compared to each other. In a study with a similar design, Hermans, van Dijk and Christoffels (2007) tested experienced sign-language interpreters (spoken Dutch – Dutch sign-language) and sign-language interpreting students. Like Liu (2001), they found no differences between the two groups on tests of general cognitive ability, including short-term memory, working memory and cognitive control, but a significant difference on the interpreting tasks. When the results of the two groups of tests (cognitive ability and interpreting) were brought together, however, a relationship was found between short-term memory and interpreting between Dutch and signed Dutch (both directions), and between working memory and cognitive control and interpreting from Dutch into sign-supported Dutch.[9] Such results indicate that while the *group* performance may not differ, when *individual* performance is taken into account, relationships exist.

3.3. Where next?

Previous empirical research into working memory in interpreters indicates that there may be a cognitive advantage of interpreters over non-interpreters in their higher cognitive abilities. While differences on simple memory tests were not unequivocally supported, evidence from complex memory tests, such as the reading span and listening span, point in the direction of a possible difference between individuals with interpreting experience and training, and those without. It is less clear whether the difference is pre-existing or due to

[9] Sign-supported Dutch (Nederlands met Gebaren, NmG), like Manually Coded English, is not a sign-language in its own right, rather a regular spoken language supported by signs. Typically, it follows the grammar and form of speech of the spoken language. It is mostly used by non-native sign-language users, such as people who became deaf later in their life (and have already acquired a spoken language as their first language), or by hearing persons who need to communicate with deaf persons.

practice, as there is no clear evidence of difference between experienced interpreters and interpreting students. Furthermore, research looking into the relationship between working memory and interpreting indicates that there is an advantage of better working memory in early stages of acquisition of the interpreting skill – both untrained bilinguals and interpreting students performed better on an interpreting task when they also demonstrated better working memory capacity. Interestingly, this relationship was present even when interpreting was assessed using fairly crude measures of interpreting – overall accuracy or quality using a very limited scale of 0-5 points for the whole task (Christoffels, 2004; Tzou et al., 2011) or a binary scale 0-1 for individual content units (Hodáková, 2009). The most obvious gap in this line of research is an investigation of the relationship between working memory and interpreting in experienced professional interpreters. Closing this gap, or rather making an attempt at reducing it, is the aim of the present study.

So far, a relationship between working memory and interpreting has been found at the highest level – working memory was related to overall interpreting performance. As we have discussed in Chapters 1 and 2, however, working memory itself is a very complex construct with a number of specific functions. Similarly, the interpreting task can be broken down into a number of specific processing steps and approached from a number of perspectives. This study therefore takes a more differentiated view of both working memory and simultaneous interpreting, and takes a number of measures of both constructs. For working memory, both its storage and central executive functions are measured. In simultaneous interpreting, several aspects of specific linguistic processes are considered, such as the processing of numbers, as well as more global measures, such as trends in ear-voice span behaviour. All the tasks are administered to a single group of professional simultaneous interpreters, and individual performance on the interpreting task in relation to working memory is assessed. The research questions this study addresses are therefore the following:

1. Is there a relationship between working memory and simultaneous interpreting performed by professional interpreters?
2. Is working memory involved in all aspects of simultaneous interpreting to the same extent, or do different functions of working memory support different processes in simultaneous interpreting?
3. How strong is the relationship between working memory and simultaneous interpreting? Do data support the notion that working memory is a crucial mechanism of simultaneous interpreting performance?

The rest of this dissertation will attempt to provide an answer to these questions. Chapter 4 describes a methodological framework for the study, the selection of working memory tests and choice of simultaneous interpreting parameters. It also includes a full, technical methodological chapter, including detailed description of all tests, scoring procedures, stimulus materials for the interpreting tasks and details of the technical equipment used. In Chapters 5 and 6 we will present the results of the data analyses and discuss them in relation to the research questions asked and in the context of previous research.

Chapter 4. Method

4.1. General methodological considerations

The study to be reported in the following chapters was designed as an exploratory descriptive quantitative study of individual differences in working memory and simultaneous interpreting, conducted with professional interpreters. Before reporting the study, it is necessary to provide a methodological context and justification of the methodological approach, choices and decisions, including inherent limitations.

First of all, this study is **exploratory**. As discussed in Chapter 3, working memory research in interpreting has so far largely focused on testing the superiority hypothesis that working memory capacity is larger in interpreters than in the general population. The few studies which tested the involvement of working memory in simultaneous interpreting (as opposed to in interpreters) have failed to establish the connection. Assuming the connection is there – and from the literature review of working memory involvement in areas such as language processing it seems a given that the connection exists – the problem becomes one of finding which components of working memory are involved and how. Since there is little previous research to guide us towards an answer to this question, the best approach seems to be to carry out a broad survey of the area. Therefore, in this study we aim for breadth rather than depth of investigation, without specific hypotheses constraining the analyses. A broad range of working memory and interpreting measures was selected to maximise the chance of discovering a clue as to the relationship between the two constructs. The findings can then serve as a basis for future, more specifically targeted and better controlled studies which will further illuminate the involvement of working memory in simultaneous interpreting.

Similarly, this study was designed as a **descriptive**, rather than an explanatory study. Given the large scope of the study, the aim is to describe what is necessarily bound to be a complex network of associations. Some relationships may be unexpected, while others may not be found despite intuitive plausibility. The broad design also means the measures are rather course and a detailed explanation may not be possible without further experimental work.

Thirdly, the approach to data analyses is **quantitative** in nature and results are derived by means of statistical analyses. This means that the choice of interpreting parameters selected for analysis was constrained to those where quantification is possible. The absence of qualitative measures obviously means a number of important aspects of the interpreting process were not taken into consideration. These include, but are not limited to, parameters such as interpreting strategies or extralinguistic information.

Given the exploratory nature of the study, the **statistical analyses** of the data too were limited to descriptive methods. While inferential statistics were carried out, it is the descriptive nature of the results that is of main interest here. The analytical methods were mostly those of correlation and regression, which allow examining the nature and strength of a relationship between two or more variables. Statistical significance (*p* value) will be reported as well, but less emphasis will be placed on it than in hypothesis-testing research. Additionally, the sample tested in this study has 28 interpreters, which is – from a statistical point of view – a low number that affects the reliability of statistical findings. Two considerations were made. We retained the traditional level of 0.05 for establishing significance. Level of significance is a numerical value which tells us how likely it would be to obtain the results we obtained, if in fact there was no relationship in the data. At 0.05, there is a 5% probability that we would come to this incorrect conclusion. Another consideration has to do with power, another numerical value, which tells us how likely it is that we find a relationship which really is there despite all the "noise" that influences our data. We have calculated the statistical power for the present study, using G*Power 3 (Faul, Erdfelder, Lang and Buchner, 2007), and have the following expectations: relationships which have a strength of r= 0.5 (Pearson correlation), in a sample of 28 interpreters and with significance probability level at 5%, have an 80% probability of being found. However, for relationships with strength at r=0.3, and the same parameters of sample size and significance level, the probability of detecting the relationship drops to 35%. There is, therefore, a very real possibility that some of the weaker relationships will not be detected. This is why the statistical analyses should be primarily considered as evidence of trends, rather than confirmatory results, and the significance probability should not be the primary factor – and definitely not the only one – in evaluating the results.

Process-oriented research in interpreting has mostly made use of two methods: **experimentation and verbal reports**. Verbal reports, such as retrospective think-aloud protocols (e.g. Ivanova, 1999), are methods advocated in the study of expertise (Ericsson and Simon, 1980) and the basic idea is that interpreters provide a piece of interpreting and afterwards they are asked to comment on what they did. The method is best suited to study the processes of conscious decision making and application of strategies (high-level processes). It does not allow, however, to study processes which are subconscious and of which interpreters are not aware (low-level processes, such as perception, much of comprehension, etc.; Ericsson and Simon, 1993).

Experimental methods, on the other hand, are based on manipulation of the testing situation by creating two or more conditions, and then comparing the performance of interpreters based on the manipulated parameters. The manipulation typically consists of comparing two different groups of interpreters, such as novices and experts (e.g. Padilla, Bajo and Padilla, 1995) or by manipulating text (e.g. genre, Liu, 2001), speaker (e.g. speed, Gerver, 1969/2002) or context (e.g. preparation, Anderson, 1994) variables. This type of design makes the assumption that interpreters within each group make up a homogenous group. However, as Lamberger-Felber (2001, 2003) has shown, this assumption may be false. Specifically, Lamberger-Felber tested differences in processing of different types of text by a group of professional interpreters. Unexpectedly, her findings have shown that on a whole range of measures (errors, EVS, lexis, etc.), the interpreters varied greatly and their performance differences did not correspond to the differences in text types. In other words, there was much individual variability in interpreting which could not be attributed to obvious external factors. This suggests that interpreters may behave very differently on the task and employ a variety of different sub-processes and that they have varying strengths and weaknesses for which they may compensate – or not – in different ways, that they may employ different strategies, apply different norms etc.

Such individual variability opens up another avenue of methodological approach to interpreting process research, namely the exploration of **individual differences**. This approach assumes that each interpreter is different, that these differences are measurable and, crucially, that they are observable in the interpreting output. The methodological approach adopted in this dissertation is to explore those differences in interpreting performance and see whether they can be related to differences in interpreters' cognitive functioning, specifically to their working memory and executive control functions.

The experimental method allows comparison of interpreters and non-interpreters. By using the extreme design, comparing individuals with an advanced skill (interpreters) with individuals with zero level of that skill (non-interpreters), differences related to interpreting may be discovered. This may help to identify differences which are crucial for interpreting: a given trait is found in interpreters, but not in non-interpreters and the trait can be hypothesised to form the basis of the interpreting skill. By testing interpreters only, a group of people who all possess such a trait will be examined and the trait will not emerge as a strong predictor. In the context of the present research, if superior working memory is indeed the cornerstone of simultaneous interpreting, then all interpreters will have superior working memory and there will be no individual differences to measure. Working memory will therefore not emerge as a **decisive trait** which underpins interpreting. This limitation, however, does not apply to individual working memory differences being related to differences in interpreting performance.

A second limitation is related to the fact that this study is based on participant variables. Neither interpreting skill, nor working memory functions can be manipulated, they can only be measured. Their coexistence does not allow any conclusions as to **causal relationship** between the two main constructs, interpreting and working memory. It will not be possible to say whether individuals with specific working memory functions develop a certain interpreting style, or whether interpreting practice affects people's working memory functioning.

4.2. Measuring working memory

In Chapter 2 we discussed the complexity of working memory. Previous research has shown that different components of working memory are involved differentially in language processing, bilingualism and skilled behaviour. This suggests that the various working memory functions (broadly, storage and processing) may play different roles in cognitive behaviour. It has also been shown that the properties of complex span task, the quintessential working memory tasks, such as the reading span, are not entirely understood and that these tasks themselves tap various resources.

In order to enhance our understanding of the ways in which working memory is employed in simultaneous interpreting, this dissertation takes a slightly different approach. Rather than using traditional complex span tasks, we identified a number of working memory functions which seem relevant for interpreting and tested those directly. Furthermore, as we discussed in Chapter 3, previous research on working memory in interpreting focused more on the *memory* component of working memory, while the *working* component has never been systematically explored. This dissertation therefore places a major emphasis on central executive functions. The following section describes the functions selected for testing. The selection was guided by two main principles. First, the working memory function has been described in working memory literature, it has been operationalised and an empirical test of the function has been developed. The function has been supported empirically, and there is a degree of plausibility, either based on related empirical work or on face validity, of the function being relevant for interpreting. Second, the package covers a broad range of working memory functions, both of its storage and its executive components. This approach is again somewhat coarse, but consistent and, in our opinion, the best for the exploratory nature of this study. A total of seven working memory functions were selected: two tests of storage (verbal memory, visuospatial memory) and five tests of executive functions (resistance to interference, resistance to automatic responses, updating, shifting, dual tasking). All functions should be measurable using tasks aimed specifically at the function of interest, and measure one function only to avoid problems with interpretation of results.

4.2.1. Memory functions

The memory function of working memory varies with stimulus material, most generally for verbal and visuospatial information. In this context, the important difference is not between presentation modality, but between the nature of the information itself. Therefore language presented either as a reading exercise (visually) or as a listening exercise (auditorily) is always verbal. Visuospatial stimuli, on the other hand, include colours, patterns, position of objects, trajectories, movement etc. Pure measure of the memory component can be taken using traditional tasks such as the digit span. It is important to note that these tasks do not include a processing component, and hence are considered to be tests of short-term memory rather than working memory. However, Engle and Conway (1998) argued, synthesising previous research by Baddeley and others, that verbal short-term memory is involved in language processing (such as long complex sentences) and Baddeley, Gathercole and Papagno (1998) provided evidence of its involvement in vocabulary acquisition. Visuospatial memory has been shown to be related to spatial semantic processing (Enlge and Conway, 1998). Also, this type of memory is the one which has been most often tested in the context of interpreting research, and is therefore of substantial interest for this dissertation. Short-term memory may be used during horizontal processing (de Groot, 1997) of items such as figures, which are largely context-independent (have to be dealt with as isolated items).

Two tests of short-term memory were selected. The letter span task is a variant of the digit span task, a classic test of *verbal memory*. Participants are presented with strings of individual letters and asked to recall them in the order presented[10]. The Corsi task (Berch, Krikorian and Huha, 1998) is a traditional test of *visuospatial memory*. Participants are

[10] The digit span task is known to vary between languages (Naveh-Benjamin and Ayres, 1986), largely due to the varying lengths of digit names in different languages, and was therefore considered inappropriate for this study.

presented with nine blocks scattered on a board. Several blocks are touched or otherwise pointed out in sequence and the task is to recall the sequence in the correct order. In both cases, a score is obtained to measure how successfully participants recalled the strings of letters /sequences of blocks.

4.2.2. Executive functions

Five executive functions were selected for testing. *Resistance to interference* is required to keep focus on the task at hand and to avoid being distracted by irrelevant stimuli. These may include environmental factors, such as irrelevant sounds and noise, or task context factors unrelated to task goal. In the context of simultaneous interpreting, the need to resist interference makes intuitive sense. Interpreters must disregard or cope with environmental factors, but also with distractions such as their own voice, which interferes with their listening to the speaker. Resistance to interference is again considered to be a fairly abstract function of the central executive and an ability to focus attention on task relevant factors. A traditional test is the Eriksen flanker task (Eriksen and Eriksen, 1974). In this task, participants are presented with a row of five stimuli and need to make a decision on the stimulus in the middle. The other four stimuli (two on each side of the target) are either consistent or inconsistent. If they are inconsistent, they must be suppressed so as not to interfere with the relevant response to the target. For example, participants may be asked to judge whether the target is a consonant or a vowel, and may be presented with the following series: HHHHH. All letters are consonants, they are consistent with the target. Another series is HHOHH. In this case, the target letter *O* is flanked on either side by letters *H,* which are inconsistent with the required response and serve as distractors. The difference in response time to the consistent and inconsistent series is taken as the cost of blocking out the interfering information.

Another inhibitory function is the *resistance to automatic (prepotent) responses*. Such responses may arise due to developed routines (automated behaviour) or to a response being triggered by the stimulus. A typical example in the interpreting context may be the avoidance of false cognates or postponement of interpreting until sufficient information is available to allow for planning (e.g. postponing committing to a particular syntactic structure to avoid problems later). Again, such resistance to prepotent responses is considered to be a general central executive function. One traditional task used to measure resistance to automatic response is the antisaccade task (Hallet, 1978, cited in Friedman and Miyake, 2004). In this task, participants are asked to look for a target on the screen. Before the target appears, an unrelated distractor flashes up in another part of the screen. If participants make the automated response and visually follow the distractor, they will miss the target. Hence to succeed in this task, they must resist the automatic response of looking at a prominent distractor in order not to miss the target stimulus. Another test of this function is the Stroop test used by Köpke and Nespoulous (2006), which was discussed in Chapter 3.

The third function is *updating*. This function requires evaluating continuous incoming information against information held in memory and then making necessary changes in the memory content in view of the task goal. It is difficult to overemphasise the resemblance to the demands of simultaneous interpreting: a continuous stream of incoming information, which needs to be held for a brief moment, while it is being processed, and then flushed to make room for new information. Complex tasks such as the *n*-back task are used to measure the ability to update the task-relevant information. In this task, participants are shown a series of stimuli, typically letters, and are asked to evaluate the currently shown stimulus with a stimulus presented several, *n*, steps back. Thus in a 1-back task, the goal is to say whether the currently presented letter is the same as the letter shown just before. In a 2-back task, the current stimulus is evaluated against a stimulus shown two letters back. The two previously shown letters must therefore be held in memory – the first one is being evaluated against the current stimulus, while the other is to be evaluated against the upcoming stimulus. By then, the first letter is dropped, and the currently presented letter moves to memory. This constant updating process is illustrated in Figure 4.1.

The top panel a) shows an example where the shaded letter T is currently being presented; the preceding letters R and S are stored in memory and T is to be compared to the stored letter R. In the next step, shown in b), S is presented and the memory set has to be updated by dropping R and including the previously presented T. The comparison is now being made between the presented S and the S stored in memory. Panel c) shows some possible scenarios. In an $n-2$ task, some $n-1$ and $n-3$ stimuli are typically included as well, serving as distractors.

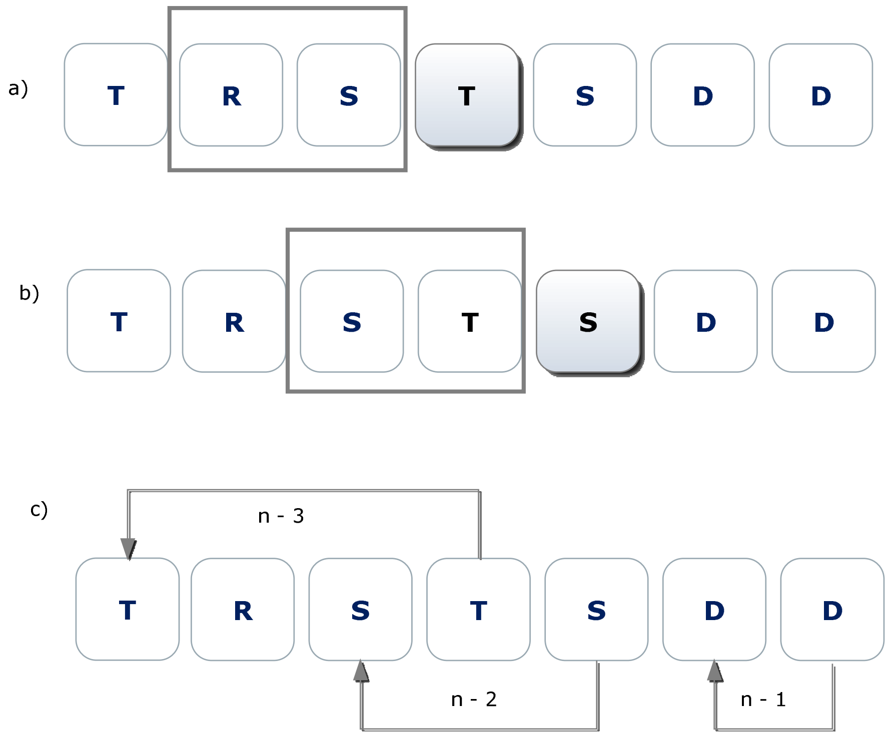

Figure 4.1 The structure of the *n* - 2 task

The fourth executive function tested is *shifting* or task switching. Miyake, Friedman, Emerson, Witzki, Howerter and Wager (2000) define this function as the ability to disengage from a current task and engage in a new one. They also point out that the exact mechanics may be more complicated and may involve other sub-functions, such as overcoming interference. The ability to shift or switch is measured by tasks such as the number-letter task (Rogers and Monsell, 1995). In this task, participants are presented with a single set of stimuli, consisting of a digit-letter pair, such as 4R, in a 2 x 2 grid. Depending on where in the grid the stimulus appears, participants make either parity judgement (odd or even) or letter judgement (consonant or vowel). The time is measured it takes to make the judgement. Some trials require two consecutive judgements of the same type, e.g. letter judgement, and hence involve the same task. Other trials are a combination of parity and letter judgement, so the participant has to mentally switch from one to the other. This is reflected in a slightly longer response time and is called the switch cost. The ability to switch between two tasks or two mental sets has again a high plausibility in the interpreting context. Interpreters process an incoming stream of auditory input and produce their own corresponding output. There is evidence that interpreters monitor their own output (e.g. Petite, 2003), and one possibility of managing the two streams is that interpreters switch between them[11].

The last function selected was *dual tasking* and attention switching, and was measured by the Barrouillet task (e.g. Lépine, Bernardin and Barrouillet, 2005). Here participants are presented with two separate tasks, a memory task and processing task, very similar to the traditional working memory tasks, such as the operation span, where the storage and processing tasks are unrelated. The Barrouillet task measures decrease in performance on the memory task due to increased attention load. By making the processing task more attention demanding, participants are prevented from switching to the memory task (refresh the memory trace), and recall of the stored stimuli tends to be impaired. The basic design of the Barrouillet task involves a memory task, such as remembering a string of letters, followed by a processing task, such as parity judgement, and a recall task of the stored string of letters. The whole task is externally paced, usually by computer, so that participants cannot complete it at their own leisure. The task is presented twice, in a slower and a faster version. The expected finding is that in the faster version, the possibility to switch attention to the memory task will be very limited, and hence the recall will be worse than in the slower version of the task. This function is closely related to attentional control, which was suggested to be crucial for interpreting (Moser-Mercer, 2005).

4.3. Measuring simultaneous interpreting

Like working memory, simultaneous interpreting is a very complex concept. Unlike working memory, the underlying processes have not been fully specified and even less so systematically studied. The issue of measuring interpreting is thus much less clear-cut. Before describing the individual measures, three major decisions that shaped the choice will be discussed: holistic versus componential approach, the choice of dependent variable and the choice of participants.

Psychologically-oriented interpreting research based on the information-processing paradigm argues for a componential approach (Shlesinger and Massaro, 1997, Frauenfelder and Schriefers, 1997, de Groot, 1997). Psychologically-oriented interpreting research rooted in interpreting studies, on the other hand, argues for a holistic approach (Setton, 2001). The

[11] An alternative is that interpreters process both streams in parallel. This issue has not been studied in detail and both possibilities remain theoretically possible. The switching hypothesis has a slight theoretical edge if interpreting is considered to be an attention demanding process. Parallel processing would be plausible if interpreting is seen as routine and automated. This issue certainly merits more attention.

starting point for both approaches is the same: interpreting is very complex behaviour. The information-processing paradigm addresses this issue by breaking the behaviour down to individual underlying processes and studying these in isolation. The advantage is that a relatively simple experimental environment allows for a much clearer understanding of what is going on. The supporters of the holistic approach argue, that such reductionist approach leads to a significant loss of the interactions among the individual processes. Since we have poor understanding of the contribution and importance of such interactions, removing them from the study design may in extreme cases lead to an investigation of behaviour which has nothing to do with interpreting. Both approaches are correct, in their own ways, and our opinion is that they are both needed. In the context of the present research, we have opted for a holistic approach with elements of information-processing design. Specifically, the testing situation was designed to resemble, as closely as possible, a real interpreting event. The texts for interpreting were based on real conference contributions and were presented in a realistic context. Interpreters were given background information, provided with a video recording of the speaker, given the space and tools they use in their normal interpreting jobs (such as taking notes, receiving contextual information, etc.; except for the possibility to consult dictionaries or team members).

On the other hand, some features of the texts were manipulated in order to achieve a certain degree of control over the stimulus materials. More specifically, we manipulated speed of delivery, text orality, topic, and also some textual and content features, such as the amount of numerical information or syntactic complexity of selected segments (see below for more details). This approach allowed us to insert some manipulated material into a fairly realistic task. We hope that it allowed for all the salient features of simultaneous interpreting as an instance of highly-skilled behaviour to be employed in the task and at the same time, provide material for an objective study of selected features.

The second major decision deals with the choice of a dependent variable. So far, psychologically-oriented interpreting research does not have established dependent variables nor design "templates" with known properties and a body of results to guide hypotheses generation and expectation of outcomes. The interpreting variables selected are therefore also rather exploratory. Three criteria were followed in the selection of variables: 1) theoretical interest for interpreting studies and some degree of intuitive justification why the measure should be related to working memory, 2) empirical feasibility, i.e. the variable can be objectively measured, and 3) each variable covers a different aspect of the interpreting process. The variables were generally divided into those providing a measure of local or global processing. Local processes were measured at specific points, while global processes were considered to span the whole task. Local processes were typically specific linguistic phenomena, while global processes have more to do with the overall task management. The following variables were selected.

Finally, a major decision concerned participants. Empirical research in interpreting studies struggles with recruiting a sufficient number of participants. Kapranov (2008) notes that most studies do not have more than ten participants. As a proxy, students are sometimes recruited. While students can often provide a good quality interpretation, it is not clear whether the underlying processes are identical to those of skilled interpreters. For this reason, we recruited only professional interpreters for the present study. “Professional” has been defined as interpreters who have passed professional accreditation test and currently actively work for the institutions of the European Union (Commission or Parliament), either as freelancers or as staff interpreters.

4.3.1. Measures of local processing

Local processing included measures of lexical, syntactic and semantic processing. *Lexical processing* was operationalised as the interpretation of figures. Unlike other lexical items, such as specific nouns, verbs or adjectives, figures offer a distinct advantage from the point of view of empirical research. They can be easily isolated both in the source text and in the target text. They are a salient feature for the interpreters and tend to be given certain prominence. Figures are typically interpreted, and an omission of a figure can, more often than not, be assumed to be an error rather than a strategic choice. Figures also lack semantic content, cannot be inferred from the context and hence offer a fairly pure measure of localised lexical processing. Additionally, they are considered to be difficult for interpreters (for the reasons mentioned), and are therefore of interest to practitioners, trainers and researchers. Because of the abstract nature of figures, we consider them to be a good candidate for the basic storage function of working memory. Upon hearing a figure, the interpreter has to store it until the previous segment has been processed, and then render it in the other language. Few other linguistic items require such reliance on memory in the context of interpreting. A plausible expectation would be that interpreters with larger storage capacity would be able to transfer more figures more accurately than interpreters with smaller storage capacity. Ten sentences with two to three figures (24 figures in total) were therefore embedded in the interpreted text and the number of correctly translated figures was counted.

Semantic processing was operationalised as sentences containing double negation. Psycholinguistic research shows that a positive affirmative clause is neutral and unmarked. Negative affirmative clauses, on the other hand, are marked, hence more difficult (Clark 1969), and their comprehension requires more neural activation (Carpenter, Just and Reichle, 2000). In the context of interpreting, Büllow-Møller (1999) has shown that interpreters make more errors in marked sentences (negative, modal, etc.) than in unmarked sentences. This suggests that negation at the sentence level (i.e. negative verb) may be more difficult for interpreters than interpretation of positive affirmative sentences. To emphasise this specific difficulty, ten sentences were constructed with a double negation of the verb. The negation was either located in two different verbs, as in *I didn't agree not to go*, or consisted of a negation of a negative verb, as in *I didn't disagree with you.* Again, all sentences were embedded in the text, and the number of correctly disambiguated negations were counted.

Finally, *syntactic processing* was operationalised as interpretation of sentences with a complex syntactic structure. Specifically, Andrews et al. (2006) have shown that working memory is associated with more successful comprehension of sentences containing relative clauses, which require the integration of several nouns and verbs into the correct relations. Similarly, King and Just (1991) have shown that participants with larger working memory capacity were better at comprehension than participants with smaller working memory capacity, and the difference was more pronounced in object-relative sentences than subject-relative sentences. Given the added difficulty of simultaneous interpreting, in comparison with self-paced monolingual reading, we opted for the simpler option. Ten sentences were embedded in the text, each with the following core structure: *subject + subject-relative clause 1 + subject-relative clause 2 + main verb*. An example of such a sentence is: *Many people (subject) who eat junk food (subject-relative clause 1) and who do not exercise (subject-relative clause2) suffer (main verb) from high cholesterol.* There is considerable distance between the subject and main verb in the sentence. In these test sentences, we measured interpreters' ability to retain the subject + main verb relation (number of correctly retained subject-verb agreements).

4.3.2. Measures of global processing

Global processing measures were those considered to span the whole interpreting task, and were related to overall performance. Three variables were chosen: vocabulary richness, planning and speed.

Vocabulary richness is a measure of how varied one's vocabulary is and how large one's mental lexicon is. Generally, individuals with larger working memory capacity acquire new words more easily both in isolation (Baddeley et al., 1998) and in context. In fact, Daneman and Green (1986) found that larger vocabulary is related to greater ability to infer meaning of unknown words from context, which they attribute to better functional capacity of working memory. Importantly, Daneman and Green tested also active production of words (rather than just comprehension) and likewise linked better performance to larger working memory capacity. In interpreting research, Lamberger-Felber (2001) has shown there is a great variability in the use of vocabulary by interpreters. Out of 2284 different words (types), only 6.6% were used by all twelve interpreters in her sample, and an astounding 45.6% were used by only one interpreter. Just over 25% of types were used by 50% of the sample. This suggests great variability in interpreters' vocabulary, which cannot be easily explained by external variables related to the interpreting task: the interpreted texts in Lamberger-Felber's study were presented under different conditions (with or without text), but the linguistic content, which presumably motivates choice of words, was the same and the effect of the conditions should not cause the variability in lexical use to be so high. Vocabulary richness was operationalised into two measures: type/token ratio and unique vocabulary. Type/token ratio is a standard measure used in corpus linguistics. It compares the total number of words in the output (tokens), with the number of unique words used (types). Unique vocabulary was measured as the number of words used only by a given interpreter.

Planning is required in interpreting to carefully balance the task and external constraints. At its simplest, interpreters cannot start interpreting as soon as they hear the first words – they need to wait for a meaningful chunk of text. A "meaningful chunk" is determined by both content and form in the sense that different languages impose different linguistic restrictions. A much cited example is the German verb located at the end of the sentence, but needed much earlier in an English sentence. In this respect, planning was operationalised as ear-voice span (EVS).

Ear-voice span (EVS), or time lag, is a temporal measure of the delay with which the interpreter articulates the output in relation to the speaker. Ear-voice span has been the subject of a number of studies, most of which focused on determining the length of the delay.

Additionally, Goldman-Eisler (1972) found that the "lag of choice" is a syntactic unit, typically noun and verb unit, and Treisman (1965) and others have shown that the lag in simultaneous interpreting is longer than in shadowing, which is thought to reflect the increased processing requirements of interpreting. Ear-voice span shows highly individual patterns, as demonstrated by Lamberger-Felber (2001). Based on an examination of interpreters' output, Lee (2002) proposed a "watershed" value of EVS: time lags longer than approximately 4 seconds are associated with increased error rates. This would suggest that interpreters need to carefully balance the distance they can keep in order to receive sufficient information to enable interpretation, and the distance that they can keep before they overload their memory and start making errors. According to Moser-Mercer, Frauenfelder, Casado and Künzli (2000), experienced interpreters have differently organised knowledge, which allows them to process the input in larger chunks, rather than individual words, resulting in longer EVS compared to students and novices. We would therefore expect that professional interpreters exhibit a certain degree of flexibility depending on the demands of the task, and suggest to measure the individual variability in EVS (individual standard deviation[12] from the individual mean EVS) as a measure of planning. If this reasoning is correct, larger standard deviation, over correctly interpreted segments, would suggest more flexibility and a better ability to plan the output. Our primary interest is again in exploring whether the ability to plan is related to individual working memory properties.

Finally, *speed* of delivery was manipulated in order to measure how interpreters cope with the varying demands. By varying speed, interpreters are presented with a different amount of input in the same amount of time. Speed is to an extent a subjective criterion. While speed of delivery in the range of 100 to 120 words per minute is considered ideal (Pöchhacker, 2004:129), factors such as familiarity with the subject, spontaneous speech versus read delivery, or intonation (Déjean le Féal, 1982) may substantially affect the perceived speed. At higher speeds, interpreters were shown to make more errors and omissions and increase their EVS (Gerver, 1969/2002). By using this measure, we wish to explore how interpreters respond to input at different speeds and whether any differences in performance can be related to differences in working memory.

All measures of simultaneous interpreting, with the exception of speed, were measured in a single text. It consisted of an eighteen-minute monologue speech, based on a real conference contribution. A total of 30 sentences were manipulated to allow for measurement of three variables: lexical, semantic and syntactical processing. The global processing

[12] A short explanation of standard deviation and other statistical concepts can be found in Annex 1.

measure of planning was measured throughout this text as well. To measure speed, two additional texts were developed. These were around 6 minutes long and contained large sections of figures and enumerative lists. The two texts were matched in terms of length, and each was presented at a different speed.

4.4. Additional measures

Apart from measures of the two main constructs of interest, additional information was collected. First of all, all interpreters provided personal information about their age, sex, interpreting training, professional experience, and acquisition of English. They also completed an English vocabulary test to provide a more objective measure of their knowledge. Since working memory is known to be strongly related to the general fluid intelligence *g*, a measure of *g* was included as well (Cattell culture fair test, Part A, Cattell and Cattell, 1950). The Cattell test consists of abstract reasoning tasks in which participants are presented with shapes arranged in matrices and have to choose one of the available options to complete the matrix. Finally, a basic choice reaction task was used. The choice reaction task consisted of a simple computerised task involving decision about the direction of an arrow presented on screen (left or right), and provided a measure of psychomotor speed. All these measures were included in the main analyses as covariates.

To summarise, the study is descriptive, exploratory, conducted with professional interpreters, and focuses on a quantitative analysis of interpreters' individual differences on two constructs: working memory and simultaneous interpreting. Working memory is measured on a series of tests targeting specific functions. These include verbal memory (measured by a letter span task), visuospatial memory (the Corsi task) and five measures of the central executive: resistance to interference from distracters (arrow flanker task), resistance to automatic responses (antisaccade task), updating (2-back task), shifting (number-letter task) and dual tasking (the Barrouillet task). The slower version of the Barrouillet task was also included as a measure of participants' complex span, which is a test similar to the traditional reading span task. This will allow comparison with previous research. Simultaneous interpreting is measured in a realistic context of a monologue conference speech, and on two additional short texts matched for content and differing in speed. Six measures of simultaneous interpreting were selected for assessment, divided into broad categories of local and global processes. Local processing included measures of lexical processing (figures), semantic processing (sentences with double negation), syntactic processing (subject-verb agreement in syntactically complex relative sentences). Global processing included vocabulary richness (type/token ratio and unique vocabulary), planning (ear-voice span) and speed (difference in performance on two texts with different speed). Additionally, the average performance on the two texts was included as a more general measure of performance under high speed of delivery. Measures of the two constructs (working memory, simultaneous interpreting) were collected and subsequently analysed using correlational and regression statistical techniques. This section described the main methodological issues and decisions in order to provide framework for the evaluation of the findings presented in Chapter 5. The following section contains a more detailed, technical method.

4.5. Specific method

4.5.1. Participants

A total of 28 participants were recruited for the study. Recruitment criteria were the following: a) the participant is a professional interpreter. Professional interpreter was operationally defined as an interpreter who is accredited to work for the institutions of the European Union, either as a staff interpreter or a freelance interpreter (auxiliary conference interpreter), and who exercises the profession actively; b) the participant's mother tongue is either Czech or Dutch; c) the participant's professional (accredited) language combination includes English. Additional information about participants was collected (age, sex, professional experience, etc.) and analysed. This information was supplementary and entered into analyses as covariates, but did not in any way constrain recruitment.

4.5.2. Apparatus

All tasks (working memory and simultaneous interpreting) were presented on a portable computer HP Compaq nc8430, with an Intel T2500 Core Duo (2.0 GHz) processor, 80 GB 5400RPM SATA hard-drive and 2GB DDR2 667 PC2-5400 memory. The computer had a 15.4-inch WSXGA+ screen with a maximum resolution 1680 x 1050, and an ATI Mobility Radeon X1600 graphic card with 256 MB of discrete video memory. The machine runs on the Microsoft Windows XP Professional platform.

Working memory tasks were programmed and presented as computer-controlled experiments using the E-Prime 2.0 software (Schneider, Eschman and Zuccolotto, 2002). All tasks were presented using a 1280x768 resolution and 32 colour bit depth. Tasks involving sounds were presented as stereo *.wav* files recorded at a sample rate of 22050Hz with a resolution of 16 bits per sample. Responses to tasks were logged using a standard keyboard in E-Prime 2.0.

Simultaneous interpreting tasks were recorded using a Sony HDV 1080i digital video camera. The recordings were then digitized using a digital video tape reader. The digital recordings were edited (picture and sound) using Microsoft Windows Movie Maker 5.1 software, and saved as *.avi* files (DVD quality video files).

Participants' performance on the simultaneous interpreting tasks was recorded using an external microphone Philips SBC MD150 (27mm wide diaphragm, uni-directional polar pattern/sensor, integrated windshield) and amplified using Bandridge Soundstage 150 audio mixer. The recording was made using Roland Edirol R-09 24 wave/mp3 recorder (MP3 format recording, MPEG-1 audio layer 3, sampling rate of 44100/48000 Hz, at 64/96/128/160/192/224/320 kbps bit rates; playback: MP3 format, MPEG-1 audio layer 3, sampling rate of 32000/44100/48000 Hz, at 64/96/128/160/192/224/256/320 kbps bit rates or variable bit rate). The source text was brought to Edirol via an in-line connection, as was the interpreters' performance. This allowed for a dual-track, with one track being the source speech and the second track being the interpreting. Participants listened to the source text from Edirol (as the audio output connection on the computer had to be used for the in-line connection between the computer and the recording device). Figure 4.2 depicts the setup.

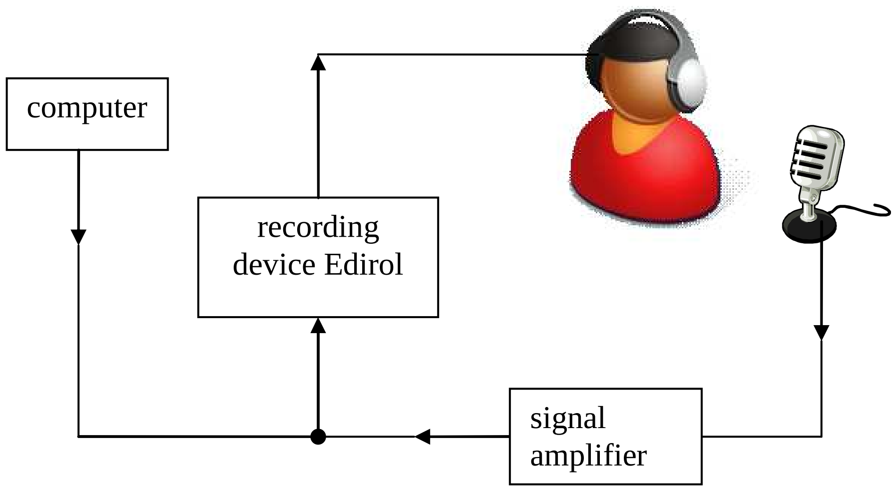

Figure 4.2 Diagram of the technical setup for simultaneous interpreting tasks

4.5.3. Materials

Demographic information

Demographic information about participants was collected. This included personal information (age, sex, highest level of education, formal interpreting training), professional profile (interpreting experience, language combinations) and English profile (self-rating of English comprehension skills, formal and informal study of English, interpreting experience related to English).

Language test

The English version of the DIALANG[13] placement test was used to provide a measure of participants' level of English. The test was re-created as a paper-and-pencil test. Participants received a sheet of paper with 50 English verbs (for example, *to campaign, to fear, to motivate*) and 25 non-words (strings of letters closely resembling regular verbs, but having no semantic meaning in English, for example *to futt, to numbelate, to driggle*). Their task was to identify the non-words. Participants had three minutes to complete the test. For each correct answer (i.e. marked non-word, unmarked verb), participants received one point; the maximum possible score was 75. It should be said that DIALANG assesses language learners up to the CEFR level C2. It was therefore expected that most participants should achieve a very high to full score, thus confirming their high level of English.

General cognitive ability

Cattel Culture Fair Test

The paper-and pencil version of Cattel Culture Fair Test Scale 3 (Cattell and Catell, 1950) was used to establish participants general cognitive abilities. Part A of the test was used, comprising 50 problems in four different sections: series (13 items), classifications (14 items), matrices (13 items) and conditions (10 items). Participants received general instructions, and prior to each section the experimenter guided them through worked examples to make sure they understood the task. Participants then had a time limit for completing the section (3 minutes for series and matrices, 4 minutes for classifications and conditions). The score was the number of correctly solved problems.

[13] DIALANG is a free language assessment software tool developed by European universities with the support of the European Commission. It allows self-assessment of language knowledge according to the Common European Framework of Reference.

Arrow test

The arrow test is basic psychomotor choice reaction task. A black arrow 10mm x 10mm appeared in the centre of a white screen, pointing either to the left or to the right. Participants were asked to indicate, as fast as possible, the direction of the arrow by using a left or right key. The test consisted of 100 experimental trials, with no practice trials and no feedback on accuracy. The task duration was approximately 2 minutes. The arrow test was completed twice by each participant, always as the last test of each session. The purpose was to obtain a measure allowing comparison of performance in the two sessions for signs of fatigue (if completed on the same day) or difference in alertness (if the two sessions were completed on different days).

The dependent variable was the mean response time. The overall accuracy rate was 98.2%. All incorrect responses and trials following incorrect responses were deleted (due to abnormally long responses), responses exceeding the personal mean by +3 standard deviations were deleted. The mean response time was calculated for each participant and each session.

Working memory tasks

Verbal span: letter span task

Series of 5-9 letters (selection from the following consonants: B, C, D, G, K, P, Q, T) were presented in a fixed (randomly established) order, with each series length presented twice (5-5-8-7-8-6-7-6-9-9). Each series began with a fixation point in the centre of the screen (a plus sign) displayed for 500ms, followed by a letter displayed for 1000ms and a blank screen for 500ms. At the end of the letter series, the English word "recall" appeared on the screen to prompt the participants to recall the letters. There was a time limit on recall – 15s, but pressing the space bar allowed to continue before the time limit was reached. If recall limit was reached, a sound signal was given. Recall was written (paper and pencil). There were two practice trials (two series of five letters), and 10 experimental trials. Scoring was done following Conway et al. (2005), using their recommended partial credit unit scoring. Each series was scored as a proportion of correctly serially recalled letters within the series, i.e. correct letters in the correct order. For example, five letters recalled correctly in a series of five letters would give a score of 1. Three letters recalled correctly in a series of six letters would give a score of .50. The overall span was then calculated as the mean of all individual series scores. The task duration was approximately 4 minutes.

Visuospatial span: Corsi task[14]

Participants were asked to memorise a sequence of highlighted squares and recall them in the order of presentation. The highlighted squares were chosen from a grid of nine squares, irregularly distributed on the screen in a fixed configuration. Sequences 3-9 squares long were presented in a fixed (randomly established) order, with each sequence length presented twice (3-6-5-4-6-8-5-9-4-3-8-7-7-9). Each sequence began with a static grid displayed for 1200ms, followed by the stimulus presentation (the same grid with one highlighted square) for 1000ms and a static grid between stimuli displayed for 500ms. At the end of the sequence, the English word “recall” appeared on the screen to prompt the participants to recall the sequence. A static grid was displayed again. During recall participants used a computer mouse to click on the previously highlighted squares in the same order in which they were presented. When they could not recall a specific item, they pressed a designated key to indicate a blank position. There were two practice trials (two sequences of three squares), and 14 experimental trials. Scoring was done following Conway et al. (2005), using their recommended partial credit unit scoring. Each sequence was scored as a proportion of correctly recalled squares within the sequence. For example, five squares recalled correctly in a sequence of five squares would give a score of 1. Three squares recalled correctly in a sequence of six squares would give a score of .50. The overall span was then calculated as the mean of all individual sequence scores. The task duration was approximately 5 minutes.

Inhibition – resistance to interference: arrow flanker task

The design of the arrow flanker task was loosely based on Fan, McCandliss, Sommer, Raz, and Posner (2002). In this task, the participant’s task was to indicate the direction of a central arrow (left or right) presented between distractors. On each side of the target arrow, there were two dashes (neutral condition) or two arrows. The distracting arrows pointed in the same as or in the opposite direction as the target arrow (facilitating or interfering conditions respectively). To respond correctly, participants had to resist interference from the distracting arrows.

An arrow appeared in the centre of the screen flanked by two arrows or lines on each side. The arrows and lines all subdued a horizontal visual angle 0.58°, and the five symbols together subdued a horizontal visual angle 3.8°. The arrows were black and appeared on a white background. The arrows appeared for a total of 1500ms, followed by a 500ms blank. Participants were asked to indicate the direction of the central arrow by using a left or right

[14] I am grateful to Arnaud Szmalec for making a computerised version of the test available to me.

key. There were three flanker conditions: congruent (all five arrows pointed either left or right), incongruent (the central arrow pointed in a direction opposite to the four flanker arrows) and neutral (the central arrow was flanked by four straight lines). The experiment consisted of 18 trials with feedback on accuracy (tone for incorrect and no responses) and 102 experimental trials. The task duration was approximately 4:30 minutes.

The dependent variable was the ratio of the mean response times in the incongruent condition and the neutral condition. Scoring generally followed Oberauer et al. (2003:174). The first trial was deleted for all participants (an outlier in most cases) together with response times shorter than 200ms, all trials with incorrect responses and trials with response times exceeding 3 standard deviations of individual means. Due to a large number of incorrect responses, all data for one participant were deleted. The response times were reciprocally transformed (1/RT), means for the incongruent and neutral conditions were calculated and converted back to response time. The (incongruent mean RT)/(neutral mean RT) ratio was determined for each participant. A lower score indicates smaller interference effect, and hence better performance at inhibiting the interference from irrelevant stimuli.

Inhibition – response inhibition: antisaccade task

The antisaccade task was modelled on Friedman and Miyake (2004). The goal was to indicate the direction of an arrow which appeared on the right or left side of the screen, preceded by a distractor appearing on the opposite side of the screen. To complete the task correctly, participants had to avoid the automatic response – visually follow the distractor, as they would miss the target item, the arrow.

Participants were seated 45cm from the screen. A fixation point (a plus sign) was displayed in the centre of the screen for a variable amount of time between 1500ms and 3500ms (in 250ms intervals – nine time periods altogether; random selection), followed by a visual cue. The cue was a 0.3cm black rectangle, appearing 8.6cm to the left or right of centre (randomly), for 175ms. After that, the target appeared on the opposite side of the screen (again 8.6cm to the left or right of centre). The target consisted of a white block arrow pointing up, left or right, enclosed in a 1.5cm white square. The target remained on-screen for 150ms, after which it was masked by grey cross-hatching, for a period of 1500ms. The participants' task was to indicate the direction of the arrow, using keyboard keys labelled left, up, right. There were 22 practice trials, followed by 90 experimental trials. The dependent variable was the proportion of correct responses. The task duration was approximately 7:15 minutes.

Updating: 2-back task[15]

A fixation point (a plus sign) was displayed in the centre of the screen for 3000ms. It was followed by the presentation of a letter for 500 ms, followed by another fixation point for 2500ms. The presentation times were fixed. Letters were presented in both upper and lower case to minimize visual memory. There was a practice list of 16 items, and three experimental lists of 48 items each. They included 2 initial items, 15 targets and 31 non-targets. These were divided into 25 neutral mismatches, and six interfering items: three were at a position 1-back and three were at a position 3-back. The dependent variable was the proportion of correctly identified letters. The task duration was approximately 10 minutes.

Dual tasking: Barrouillet task

This task was modeled on Lépine, Bernardin and Barrouillet (2005). In this task, participants were asked to remember and later recall a string of letters. Between the presentation of the letters and their recall, participants completed a processing task designed to prevent them from switching attention to the memory trace and refresh it. When the processing task is fast, participants exhibit worse recall than when it is slow, regardless of the delay between presentation and recall or complexity of the processing task.

The experiment had two parts: a slow and a fast version. In both, participants were asked to memorise a series of letters and recall them in the order of presentation. The letters were eight consonants (B, C, D, G, K, P, Q, T) which have a monosyllabic name in Czech, and Dutch. Series of 5-8 letters were presented in a fixed (randomly established) order with each series length presented twice(5-7-6-6-7-8-5-8). Each series began with a fixation point in the centre of the screen (a plus sign) displayed for 500ms, followed by a letter displayed for 1000ms. Each letter series was then followed by a processing task designed to capture attention and prevent attention switching to rehearsal. At the end of the letter series, the English word “recall” appeared on the screen to prompt the participants to recall the letters. There was a time limit on recall – 15s, pressing space allowed to continue before time limit was reached. If recall limit was reached, a sound signal was given. Recall was written (paper and pencil).

The processing task consisted of a parity judgement task. Participants were presented with a series of eight digits, presented one by one, and indicated by pressing a key whether the digit was odd or even. Each series consisted of four odd and four even digits (2-9). The digits were selected at random and could appear maximum twice in one series. The

[15] I am grateful to Arnaud Szmalec for making stimuli material for this test available to me.

processing task had two conditions: fast and slow. In the slow condition, the digit appeared on screen for 1125ms and was followed by a delay of 375ms, while in the fast condition, the digit was displayed for 600ms and was followed by a delay of 200ms. Participants thus had 1500ms to make a judgement in the slow condition, and 800ms in the fast condition. The processing time between two successive letter presentations was then 12s in the slow condition, and 6400ms in the fast condition. There were two practice rounds, one for each speed condition. For the parity judgement task, participants were given five strings of eight digits each, with feedback (in case of an incorrect or no response, a short sound signal was given). This practice was criterion based (80% accuracy rate). Then the whole task was practiced - two series of five letters, with no feedback. Recall time limit in the practice round was 10s. The whole task sequence was thus: slow version – parity practice, task practice, experimental task, fast version – parity practice, task practice, experimental task. The duration of the task was approximately 15 minutes. The dependent variable was the ratio of the fast condition score to the slow condition score. Scoring was done following Conway et al. (2005), using their recommended partial credit unit scoring. Each series was scored as a proportion of correctly recalled letters within the series. For example, five letters recalled correctly in a string of five letters, would give a score of 1. Three letters recalled correctly in a string of six letters would give a score of .50. The span for each condition (slow, fast) was then calculated as the mean of all individual series scores.

Complex span task

The slow condition score of the Barrouillet task was used as a measure of participants' working memory capacity.

Shifting: number-letter task

The task was modelled on Rogers and Monsell (1995) and Miyake, Friedman, Emerson, Witzki, Howerter, and Wager (2000). Participants were presented with a two-by-two grid, each cell being a square sized 5x5cm. A number-letter pair (e.g. 7R) appeared in the centre of a cell. Participants performed one of two judgement tasks, depending on the location of the pair. For all pairs appearing in the top two quadrants, participants decided on the parity of the number (odd-even). For all pairs appearing in the bottom two quadrants, participants decided whether the letter was a vowel or a consonant. There were two keys available for a response, a left key and a right key at the far ends of the keyboard, each key serving a double function (one response for the letter and the number task each). The experiment was presented in three blocks. In the first block, participants performed 32 number judgement trials (plus 10 practice trials). In the second, participants performed 32 letter judgement trials (plus 10 practice trials). In the third block, the number-letter pair appeared in all four quadrants, changing location clockwise, starting from the top right quadrant. This way, participants completed two number judgement trials followed by two letter judgement trials. This lead to a regular alternation of task-switching and task-non-switching trials. There were 128 trials in this block altogether (plus 12 practice trials). There was a total of eight digits (2-9), and a total of eight capital letters (G,K,M,R for consonants, A,E,I,U for vowels). The number-letter pair was formed by one letter and one number, each selected randomly from the list. The pair appeared on the screen until a response key was pressed, or for a total duration of 5s if no response was selected, followed by a 150ms pause and a new pair. To reduce errors and longer response times due to memory load, key labels (odd, even, vowel, consonant) were present on-screen throughout the whole experiment (practice and experimental trials). During practice trials, incorrect or no responses were indicated by a short auditory signal. The dependent variable was the switch cost, calculated as the difference between the median response times in the switch and non-switch trials in the third block. Scoring was done following Oberauer et al. (2003): incorrect responses and response times shorter than 200ms were deleted. The task duration was approximately 5:30 minutes.

Simultaneous interpreting measures

Text selection and manipulation

Three texts were developed for the simultaneous interpreting tasks. Text 1 (Amnesty; see Annex 2) was a genuine conference contribution available on the internet. The seminar, entitled “Business and Human Rights Seminar” took place in London in December, 2005[16]. It was a high-level event with participants representing large corporations (such as BP, Gap), major international organisations (such as the United Nations or Amnesty International), and was the 3rd event in a series of seminars on ethical issues in international business. Text 1 was delivered by a representative of Amnesty International (text Amnesty). Text 1 was slightly shortened so as to be approximately 20 minutes long when delivered at a moderate pace.

A total of 30 sentences in the text were manipulated to provide controlled material for the dependent variables. The sentences were of three types, containing a) syntactically complex structure, b) semantic complexity, and c) numbers. All thirty sentences were

[16] http://www.bhrseminar.org/

embedded in the Amnesty text naturally fitting into the context, with the constraint that no two manipulated sentences can follow immediately one after another.

Ten sentences had a complex syntactical structure consisting of *subject + subject extracted relative clause 1 + subject-extracted relative clause 2 + main verb + verb complements.* The sentences were developed in English and then translated into the target languages Czech and Dutch (Annex 5: Manipulated sentences in text Amnesty). The purpose of the translation was to test whether the target languages have a theoretical linguistic capacity a) to express the source text syntactic structure, and b) to place similar production demand on the interpreter. Both languages have formal means for creating the same syntactic structure as the source text, which was verified by the possibility to reproduce the structure in translation and by having the plausibility qualitatively confirmed by a language professional and native speaker of each language. Production demands were measured by comparing the number of words needed to express the same idea, particularly in the segment of the two relative clauses separating the subject and main verb and for the whole sentence. The mean length of each segment in English, Czech and Dutch, together with comparison of the Czech and Dutch translation is shown in Table 4.1. The translation is, of course, purely hypothetical, and it was not possible to know beforehand what translation would actually be produced by the interpreters, but this procedure ensured a minimum level of congruency in the testing materials and provided a theoretical measure of an approximately similar task for speakers of two different languages.

Ten sentences were manipulated to contain a complex semantic phenomenon consisting of a double negation. Five sentences contained the structure *verb + negation (not) + verb + negation (not)*, as in *We did not decide not to go*. Five sentences contained the structure *verb + negation (not) + negative verb,* as in *We did not disagree*. As in the case of syntactically complex sentences, the stimulus material was first produced in English, then translated into Czech and Dutch to verify linguistic viability of the material and the approximate production demands, measured in the number of words required to express the same idea in the target language. This information is shown in Table 4.1.

Finally, ten sentences were manipulated to contain two or three figures. The sentences were developed in English, translated into Czech and Dutch, and the target language versions compared for the overall sentence length (in words) and length of the numbers (in syllables). Table 4.1 provides an overview.

Table 4.1 Characteristics of the manipulated sentences in the Amnesty text

	English	Dutch	Czech	paired samples t-test Dutch- Czech (p value)
Syntax: mean lengths (in words)				
- sentence	31.6	30.7	29.4	.11
- relative clause 1	6.7	6.6	6.3	.58
- relative clause 2	7.4	6.9	6.1	.04[a]
- distance between subject and main verb	14.1	13.5	12.4	.11
Semantics: mean length (in words)				
- sentence	12.6	12.0	10.1	.11
Figures: mean length				
- sentence (words)	19.0	19.1	17.6	.11
- figures (syllables)	4.25	4.7	4.5	.21

[a]During the materials development, an error was made in one of the t-test parameter specifications, which lead to the conclusion that there is no difference between the relative clauses. The error was discovered too late to change the testing materials. Crucially, the distance between subject and main verb remains comparable in both languages.

Another text was selected from a background material to another contribution from the same event. It was a written country report on how companies in a given country comply with human rights. The text was significantly shortened so as to be approximately 5-6 minutes long when delivered at a moderate pace. The short version served as a basis for the development of two other texts, each being the said country report on a different country, Brazil and China (Annex 3). Each of the two texts contained an identical introduction and conclusion. The main body of the text consisted of a) a list of industry sectors included in the report and the number of companies analysed in each sector, and b) a list of various human rights and the number of companies which support the specific right. The lists were either presented as a list of items, or embedded in full sentences. Where the list was embedded, the text providing context was identical in the two texts, so that the only difference between the two texts was in the two lists.

Lists for each text were then matched for length of the original delivery (English) and length of the translation into Czech and Dutch (Annex 6). A list of industry sectors and human rights was drafted in English. The length of each was measured in the number of words, for English and the two target languages. For example, *forestry and fishing* is three words long in English, and three words long in both Dutch (*bosbouw en visserij*) and Czech (*lesnictví a rybářství*). The same procedure was used for the human rights names. The list was then divided in two parts, each of which was embedded in one company text, in a way so that the length of the manipulated segments was the same in both texts in English, and that it required roughly the same number of words in their translation into Czech and Dutch. A similar procedure was used for the numerical information, with one difference. Numbers do not carry semantic information, cannot be stored as concepts and must be therefore remembered verbatim. In that case, it is important to match not only the number of words, but also word length. The lists of figures were therefore matched in terms of syllables rather than words, in all three languages, i.e. the two texts in English were matched, and each required a translation of similar length into Czech and Dutch. Of course, in both cases, it was not possible to predict what translation the interpreters would choose, but this procedure ensured that both target languages allow translation of a given length. Quantitative characteristics of both texts are shown in Table 4.2 and Table 4.3.

Table 4.2 Properties of the manipulated lists (concepts, figures) in the China and Brazil texts

Concepts	Text	Number of items	Mean length (number of words)		
			English	Dutch	Czech
Human rights	China	8	4.6	4.6	4.4
	Brazil	8	4.1	4.0	3.6
Industries	China	8	2.4	2.3	2.5
	Brazil	8	2.6	2.3	2.5
Figures			Mean length (number of syllables)		
			English	Dutch	Czech
Human rights (set 1)	China	9	4.6	5.1	4.6
	Brazil	9	4.1	4.6	4.2
Human rights (set 2)	China	8	5.1	7.0	6.8
	Brazil	8	4.4	5.9	6.1
Industries	China	10	2.6	2.6	2.8
	Brazil	10	2.7	3.0	2.8

Table 4.3 Comparison of the manipulated lists in the China and Brazil texts

	Paired samples t-test – *p* value				
	Concepts		Figures		
	Human rights	Industries	Human rights set 1	Human rights set 2	Industries
Text Brazil – Text China					
- English – English	.53	.35	.38	.20	.80
- Dutch – Dutch	.45	1.00	.18	.16	.57
- Czech – Czech	.34	1.00	.58	.48	1.00
Text Brazil					
- Czech – Dutch	.35	.17	.40	.45	.34
Text China					
- Czech – Dutch	.17	.17	.14	.35	.56

Video and audio recordings

All three texts, Amnesty, China and Brazil, were recorded by a native British English male speaker with a neutral accent. The speaker was given the texts beforehand in order to be able to familiarise himself with the text, and was instructed to deliver the manipulated segments especially carefully (clear pronunciation of the negative particles in the semantic manipulations and of figures in the sentences with numbers). All speeches were written and read, with no attempts to oralise them, in order to make the stimulus material challenging for even the most experienced interpreters, and to avoid ceiling effects (perfect or near-perfect scores indicating the task was too easy).

The Amnesty text was recorded at two speeds, 125 words per minute (wpm) and 145 wpm. The two versions were piloted with two interpreters, and the slower version was selected for testing, as it was considered to be sufficiently difficult. Similarly, the China and Brazil texts were recorded at two different speeds each. The faster version of the China text (138 wpm) and the slower version of the Brazil text were used (117 wpm), as this combination provided the largest difference in the average speed of delivery.

In all recordings, the speaker was seated against a white background at a table, and his head and torso appeared in the picture. The recording allowed a good visual perception of the speaker's face and facial movements, including lip movements and hand gestures.

While every effort was made to ensure maximum video and audio quality, there were some natural limitations of the technology used. Specifically, several participants commented on the quality of sound, which was inferior to the standards they are used to in their professional environment, although not to the extent that it would seriously hamper their performance. Prior to testing, a sample of the recording was shown to three professional interpreters/researchers (none of whom participated in the study) to verify the recording quality and suitability for laboratory testing under environmental conditions simulating as closely as possible a real interpreting event.

Dependent variables

Syntactic processing: The ten manipulated sentences containing a complex syntactic structure served as a measure of syntactic processing. For each sentence, interpretation was assessed as either preserving the subject-main verb agreement across the two intervening relative clauses, or not preserving the agreement. Accuracy and completeness of the rest of the sentence was not evaluated in any way. The maximum possible score was 10.

Semantic processing: Disambiguation of the double negation was assessed as either correct or incorrect. The disambiguation could have been achieved by similar grammatical means (using negation), or by an alternative way of expression. For example, the sentence *Some companies do not respect the rule not to employ children* was interpreted as *Some* *companies do not respect the ban on child labour*. The grammatical composition of the sentence is different, but the semantic complexity was correctly disambiguated. For each correctly disambiguated sentence, one point was awarded. The maximum possible score was 10.

Lexical processing: The ten manipulated sentences contained a total of 24 figures. Each figure was scored as either correct or incorrect. A figure was scored as correct if it had been interpreted with complete accuracy. Approximations or rounding were not accepted. The maximum possible score was 24.

Vocabulary richness: This analysis requires a direct comparison of lexical units in the target language. Therefore, only data from the 20 interpreters working into Czech were used, as it was not possible to directly compare lexical items in Czech and Dutch simultaneously. A segment of text was selected from the middle of the text. The segment was 374 words long and encompassed the section on slavery, child labour and technology from the Amnesty text. Transcriptions of the interpretation were loaded into AntConc, corpus management software (Anthony, 2011) one by one, and individual word lists were compiled. Each word list was exported to Microsoft Excel and cleaned: all numbers and numerals were deleted, as were morphological forms (declensions and conjugations) of the same word (*do – did – done*), negative forms of verbs, comparative and superlative forms of adjectives, other than personal pronouns, and all slips of the tongue and unfinished words. Mispronounced words were restored to their correct form. The clean list contained all types (unique words) used by an interpreter. Individual vocabulary richness was then calculated using the standard measure of type/token ratio, i.e. the number of unique words divided by the total number of words.

As a second measure, a personal unique vocabulary score was determined as the number of words used by one interpreter only. For this analysis, all individual word lists were compiled. For each interpreter, the number of words used only by that interpreter was counted.

Planning: Ear-voice span (EVS), the distance the interpreter keeps from the speaker, or temporal delay between source text and target text, was selected as a measure of planning. EVS was measured at the beginning of the 30 manipulated sentences. The double-track recordings for each participant were loaded into Adobe Audition (Adobe Systems Incorporated, 2003). The screen was set to grid unit of 2s, with approximately 40s of audio file visible on a screen to achieve uniform measurement (for details and comparison of various ways to measure EVS, see Timarová, Dragsted and Hansen, 2011). A cue range was marked in the file, with the cue beginning set at the measurement point in the source text (the beginning of a manipulated sentence), and the cue end set at the measurement point in the target text (the beginning of a manipulated sentence). The measurement was made on the basis of semantic correspondence. For example, a sentence in the target text may have been truncated (contain only some of the source text information) or it may have formally started, but contain a large gap, as in *And as mentioned before… (2s pause)… if a company does not respect the ban on child labour…* . In such a case *and as mentioned before* may be a norm-induced filler ("keep talking"), which is not semantically motivated by the source text. The measurement would then be made on *if a company*. Missing values correspond to sentences, which were omitted in their entirety. The distance between the two cue points (cue range length) was calculated by Adobe Audition. This resulted in 30 individual values for each participant. Due to large variability in the length of the EVS, median EVS was calculated as a measure of the average time lag. Standard deviation of individual time lag (calculated on raw data) was calculated as a measure of variability in individual EVS.

Effect of speed delivery: The effect of speed delivery was measured as the difference in the number of correctly interpreted items (companies-difference) between two matched texts, a fast text (China) and a slow text (Brazil). Each text contained a total of 72 items[17], either as figures or as lists of industry sectors and various human rights. Each participant interpreted both texts in a fixed order (China, Brazil). The total of correctly interpreted items in each text was counted. The scoring procedure applied was very strict and in line with scoring recall in memory tests (see details of scoring for the letter span, Corsi or Barrouillet tasks). Each item had to be interpreted fully and accurately, approximations or partial interpretation was not accepted. For example, if *food and beverage industry* was interpreted as *food industry*, the item was assessed as incorrect. Similarly, *45.3%* interpreted as *45 %*, *45.4%* or *around 45%* were considered incorrect. The maximum possible score for each text was 72. Additionally, the average of correctly interpreted items in the two texts (companies-average) was taken as a measure of accuracy in conditions of high speed of delivery.

4.5.4. Procedure

Interpreters were recruited via personal contact or email by the researcher. An introductory information sheet was prepared which explained the nature and basic aims of the study, and which described in broad terms what they would be expected to do. The introductory document can be found in Annex 7.

[17] Some list items were repeated in the text.

Individual appointments were made with interpreters for the testing. Depending on the participants' availability, the testing was carried out in one or two sessions. The tests were divided into two blocks of approximately 90 minutes each. In case of a single session, participants completed Block 1, had a one-hour break and then completed Block 2. The entire session lasted approximately 4.5 hours, including breaks and time necessary for setup of equipment etc. In case of two sessions, participants completed Block 1 in the first testing session and Block 2 in the second testing session. Each testing session lasted approximately 1 hour 45 minutes, including breaks and time necessary for setup of equipment etc. Block 1 and Block 2 had a fixed order of tests and all participants completed the tests in the same order. The testing location was determined by participants' availability: some participants were tested in their homes, others were tested in interpreting booths at their place of work.

Block 1 consisted of an introduction, where the basic goals of the study were reiterated. Participants then signed a consent form (Annex 8) and completed a demographic information sheet (Annex 9). Then they completed the English vocabulary test (Annex 10), letter span task, arrow flanker, Corsi task and number-letter task. A short 5min break was followed by the antisaccade task and an interpreting task (text Amnesty). For the interpreting task, participants were given basic contextual information about the event (seminar programme) and shown a video recording of an introduction to the event (Annex 4). Finally, interpreters completed the arrow test. Block 2 consisted of the Cattell Culture Fair test, Barrouillet task, followed by a 5min break. Participants then completed the 2-back task, and did another interpreting task (texts China and Brazil). At the end of Block 2, participants completed the arrow test again. They were given an opportunity to review their consent form and ask any additional questions.

Chapter 5. Results and discussion

In this chapter, we will present the analyses and discussion of the results. The chapter is divided into four main sections. Section 5.1 presents descriptive and exploratory data analyses, including participant characteristics and basic quantitative description of the tests administered. Section 5.2 focuses on the structure of working memory and Section 5.3 on the structure of simultaneous interpreting performance. Finally, in Section 5.4 we will present correlational analysis of the relationship between working memory and interpreting, and regression analyses of selected complex relationships between more than two tasks.

5.1. Descriptive statistics

5.1.1. Participants

Participants in this study were recruited on the basis of three criteria: 1. they are professional interpreters, where “professional interpreters” were defined as individuals who have passed the EU inter-institutional professional accreditation test for interpreters and currently work for the institutions of the European Union, 2. they are active as interpreters, 3. their language combination includes English as a passive language (a language they interpret from) and Czech or Dutch as mother tongue. A total of 28 interpreters were tested.

Each participant also provided information (either in a self-report questionnaire, or by means of a formal test) on their sex, age, interpreter training, exposure to and knowledge of English, professional interpreting experience, and general cognitive abilities. Summary participant statistics are presented in Table 5.1.

The participating group consisted of 18 females and 10 males. There were 20 interpreters with Czech as their mother tongue (15 females) and 8 interpreters with Dutch as their mother tongue (3 females). The mean age of the participants was 37.1 years (SD=8.2 years[18]), ranging from 25 to 55 years. All participants completed university-level education beyond bachelor level, i.e. achieved a degree which formally takes four or more years of education. Twenty-three participants were formally trained as interpreters, 5 participants had no formal training[19]. All participants were active interpreters at the time of testing and interpreting was their main professional activity, either as staff interpreters at one of the EU

[18] A short explanation of standard deviation and other statistical concepts can be found in Annex 1.
[19] Eligibility criteria to sit the EU inter-institutional interpreting test allow applications from individuals with formal training and no experience, or from individuals without formal training but proven professional experience.

institutions (European Commission, European Parliament) or as freelance interpreters for the same institutions (and possible further activity on the private market). Professional interpreting experience ranged from one to 25 years, with M=11.9 years (SD=6.9 years). Since professional activity in a year varies, participants were also asked to estimate the number of days they work per each year of their professional career, which were added up to provide an estimate of the total number of days worked. The mean professional experience in number of days was M=1457 days (SD=1075 days).[20]

Table 5.1 Participant descriptive statistics

Participants	N=28	
Sex	18 females	10 males
Age	M=37.1 years SD=8.2 years	range 25 – 55 years
Mother tongue	20 Czech	8 Dutch
Interpreting experience (years)	M=11.9 years SD=6.9 years	range 1-25 years
Interpreting experience (days)	M=1457 days SD=1075 days	range 60-4500 days
University education 3+ years	N=28	
Formal interpreter training	23 trained	5 untrained
English comprehension (subjective rating; scale 1-10)	M=9.2 SD=1.1	range 5-10
English vocabulary test (max 75)	M=71.1 SD=4.0	range 62-75

5.1.2. Knowledge of English and English as a working language

Participants' mean subjective rating of English comprehension was M=9.2 (SD=1.1). The mean age of English acquisition (i.e. when participants first started learning English, either formally or informally) was M=11 years (SD=3 years). The mean score on the English vocabulary test was M=71.1 (SD=4.0). Twenty-five participants interpret from English every time they work. The mean estimated proportion of working time that participants interpret from English is M=70% (SD=18%). Twenty interpreters consider English their preferred relay language.[21] The mean number of working languages was M=3.0 (SD=1.0). When

[20] Interpreters in this sample who had no formal training in interpreting were among the most experienced ones in this sample, with the mean number of years M=19.0 years (SD=4.2 years).
[21] The EU environment is highly multilingual; up to 22 different languages are spoken in meetings. Where an interpreter does not work from one of the languages spoken on the floor, e.g. Hungarian, she uses relay

working languages were ordered from strongest to weakest, the mean ranking of English among participants was M=1.4 (SD=0.6), i.e. for most participants, English was their first, strongest working language.

5.1.3. Differences due to sex and mother tongue

A series of two-way ANOVAs was conducted to examine relationship of sex and mother tongue to age, working experience (both in years and days) and knowledge of English as measured by the vocabulary test. All main effects of sex and mother tongue and all interactions were non-significant, indicating that there was no difference between men and women, and between Czech and Dutch speakers and sex and mother tongue will therefore not be considered as confounding variables in relation to participant characteristics.

5.1.4. Working memory measures

Data were collected and scored in line with the procedures described Chapter 4. Outliers were identified and either deleted or their influence mitigated through data transformation. To maximise the use of available data, problematic cases were deleted on a test-by-test basis, which led to variable sample size per test. Descriptive statistics for the working memory tests and the two general cognitive tests (Cattell, arrow test) are in Table 5.2. The table shows mean values and standard deviation, the minimum and maximum scores achieved, skewness, kurtosis and reliability calculated using the split-half (odd-even) method and Spearman-Brown coefficient. Reliability of all tests was generally good, with the exception of the Corsi task, which may be due to the low number of observations.

---

interpretation, i.e. uses interpreting into a known language as the source for her own interpretation. Main relay languages are English, French and German.

Table 5.2 Descriptive statistics for working memory tasks

Measure	N	*M*	*SD*	Range	Skewness	Kurtosis	Reliability [a]
Letter span (proportion correct)	28	.63	.13	.40 to .92	.36	-.38	.73
Corsi (proportion correct)	28	.70	.09	.52 to .85	-.62	-.42	.45
2-back (proportion correct)	27	.89	.06	.72 to .97	-1.04	.98	.98
Antisaccade (proportion correct)	28	.77	.14	.52 to 1.00	-.06	-.95	.93
Barrouillet (fast vs slow recall)	26	.98	.16	.67 to1.3	-.26	-.50	.83
Arrow flanker (interference effect)	27	1.07	.05	1.01 to 1.18	.64	-.60	.99
Number-letter (switch cost, ms)	27	516	254	107 to 1208	.73	1.05	.77
Complex span (proportion correct)	28	.62	.18	.17 to .90	-.78	.48	.83
Cattell (total correct)	28	28.3	4.46	17 to 36	-.46	.00	-
Arrow test (speed)	28	406	34	355 to 482	.43	-.63	.98

[a] Reliability was calculated using the Spearman-Brown coefficient and split-half (odd-even) method.

5.1.5. Simultaneous interpreting measures

Data were collected and scored in line with the procedures described in Chapter 4. Descriptive statistics for the simultaneous interpreting tests are in Table 5.3. The table shows mean values and standard deviation, the minimum and maximum scores achieved, the skewness and kurtosis of the sample distribution. For vocabulary measures, only data from the Czech interpreters were used, resulting in a smaller sample size. Recording of the two short texts (China, Brazil) was faulty and could not be scored for two interpreters. Due to the score being based on a very limited number of items or a single score, reliability was not calculated for the interpreting measures.

Table 5.3 Descriptive statistics for simultaneous interpreting tasks

Measure	N	*M*	*SD*	Range	Skewness	Kurtosis
Figures (number correct, max 24)	28	14.3	4.5	5 to 22	-.01	-.64
Syntax (number correct, max 10)	28	6.1	1.9	3 to 10	-.03	-.49
Negatives (number correct, max 10)	28	7.3	2.1	3 to 10	-.41	-.97
Median ear-voice span (seconds)	28	3.2	.8	1.95 to 4.91	.49	-.24
Ear-voice span variation (seconds)	28	1.7	.5	.96 to 2.92	.89	.33
Vocabulary richness (type/token ratio)	20	.55	.04	.46 to .60	-.66	-.08
Unique vocabulary (number of words)	20	19.0	5.5	10 to 29	.38	-.77
Companies-difference (number of interpreted items)	26	8.8	6.5	-3 to 22	.02	-.15
Companies average (number of interpreted items, max 72)	26	51.8	9.7	29 to 69.5	-.57	.06

5.2. Structure of interpreters' working memory

The first analysis focused on the structure of interpreters' working memory. Table 5.4 shows the correlation matrix of the working memory tasks. Only two correlations are significant. The letter span task and the complex span task are positively related. Both the letter span and the complex span measure the ability to store and recall presented stimuli, with the difference that the letter span task is a simple storage task, while the complex span task contains a processing task and is thus considered a better measure of working memory. The positive relationship can thus be interpreted as reflecting the common component of both tasks, the storage function of working memory. The lack of significant correlations and generally fairly low correlation coefficients support the assumption that each of the tasks selected measures a distinct function of working memory.

Next, we will examine the relationship between working memory functions and interpreters' general participant characteristics. A series of independent samples t-tests was conducted to examine differences between males and females. Males performed significantly better than females on the antisaccade task (automatic response inhibition).[22] On all other tests of working memory and the two general cognitive ability tests, there were no differences between male and female interpreters. Other participant characteristics (age, experience,

[22] M_{males}=.87, SD_{males}=.13, M_{females}=.73, SD_{females}=.12, $t(25) = -2.80$, p=.01

general cognitive ability and knowledge of English) and their relationship to working memory functions are shown in Table 5.5.

Table 5.4. Correlation matrix (Spearman) of working memory tasks

	Letter	Corsi	Complex span	Antisaccade	Arrow flanker	Barrouillet	2-back
Corsi	.133						
Complex span	.686*	.147					
Antisaccade	.063	.305	.234				
Arrow flanker	-.211	.022	.128	.214			
Barrouillet	-.056	-.119	-.226	.058	-.069		
2-back	.185	.006	.283	.057	-.030	.098	
Number-letter	-.180	-.203	.103	-.109	.056	-.106	-.372

*p < .05

Table 5.5 Correlation matrix (Spearman) of working memory tasks and participant characteristics

	Age	Experience years	Experience days	English test	Arrow test	Cattell
Experience years	.833**					
Experience days	.695**	.893**				
English test	.193	.339*	.343*			
Arrow test[a]	.621**	.399**	.337*	-.265		
Cattell	-.573**	-.575**	-.466**	-.037	-.372*	
Letter span	-.260	-.198	-.219	.100	-.149	.275
Corsi	-.378**	-.300	-.252	.535**	-.655**	.316
Complex span	-.336*	-.372*	-.381**	.110	-.277	.357*
Antisaccade	-.427**	-.377*	-.336*	.017	-.481**	.267
Arrow flanker[a]	-.318	-.503**	-.602**	-.264	-.138	-.029
Barrouillet	-.121	.117	.101	-.078	-.063	.038
2-back	-.128	-.182	-.055	-.025	-.203	.330*
Number-letter[a]	.287	.138	-.027	-.166	.302	.046

**p<.05, * .05 < p < .10
[a] lower value indicates better performance

The first two rows of Table 5.5 show interpreters' experience and its relationship to age. Both measures of experience are significantly positively related to age, which is not surprising: more experienced interpreters will tend to be older than less experienced interpreters. The very strong, but not perfect, correlation between experience measured in years and experience measured in the number of days worked is a useful reminder that interpreters with the same experience in the number of years may differ on the number of days they worked. Closer examination of the variation in the data is instructive.

Table 5.6 Partial correlations (Pearson on ranked data) controlling for age

	Experience years	Experience days	Arrow test	Cattell
Arrow test	-.273	-.169		
Cattell	-.217	-.115	-.026	
Letter span	.035	-.056	.016	.159
Corsi	.028	.016	-.580*	.131
Complex span	-.176	-.218	-.093	.213
Antisaccade	-.049	-.061	-.300	.046
Arrow flanker	-.445*	-.550*	.071	-.284
Barrouillet	.403*	.268	.020	-.050
2-back	-.144	.046	-.170	.311
Number-letter	-.191	-.321	.176	.253

*$p<.05$

Figure 5.1 shows the two measures of experience plotted against each other. The scatterplot shows that at low levels of experience, interpreters do not vary too much. However, with increased experience (measured traditionally in years), it does matter how many days interpreters work per year. Two points in the graph have been enlarged to emphasise the point (two large black dots at the level of 20 years of experience). Both interpreters have 20 years of experience, but one has worked approximately 2100 days over her career, while the other approximately 3200. In other words, if two interpreters work 100 and 70 days per year respectively, the difference may not be too great after five years (500 vs 350 days), but after 20 years, it will be substantial (2000 vs 1400 days). No study has been conducted to investigate the effect of *frequency* with which a skill is practiced on the attained interpreting skill level, and further interpretation is not possible. The data presented here do, however, illustrate the well known problem of estimating interpreters' experience level and point to the difficulties with participant characteristics, especially where studies focus on comparison of more and less experienced interpreters. It may be recommendable that at least at higher levels of professional experience where the variation makes most difference, interpreting experience is estimated using also other criteria than just the number of years.

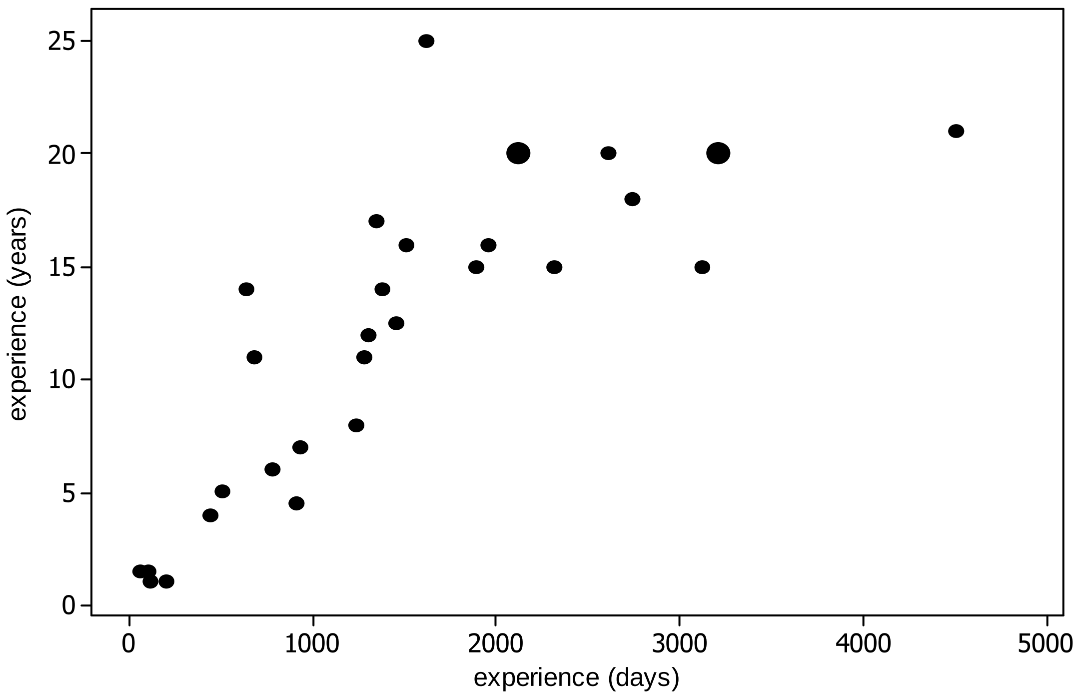

Figure 5.1 Scatterplot of two measures of interpreting experience

Returning to Table 5.5, knowledge of English is related to experience, but not to age, which suggests that this may be an interpreting-skill related knowledge, and it will be looked at in closer detail later. As for the two general cognitive ability tests, both are related to age and interpreting experience. The correlation of age and experience is positive with the arrow test, and negative with the Cattell test. The arrow test is scored in terms of reaction times, and the positive correlation means that older/more experienced interpreters are slower. The Cattell test score is based on the number of correctly solved abstract problems and therefore shows that older/more experienced interpreters perform worse (solve fewer abstract problems correctly) than younger/less experienced interpreters. Both sets of correlations therefore support the same premise that cognitive functioning deteriorates with age[23]. The fact that age and interpreting experience are themselves correlated most likely means that at least some of the relationship between the arrow test and the Cattell test on the one hand, and interpreting experience on the other hand, may be a hidden influence of age. To assess the relationship between interpreting experience and cognitive functions without the interfering influence of age, additional analyses are necessary. Table 5.6 shows the relationship between experience,

[23] This does not necessarily mean that older interpreters in our sample are cognitively "old", only that there is a general trend of cognitive patterns in relation to age. Even if older participants had scores lower by one point only, or took 1ms longer to respond, this would show as a statistically significant relationship, if sufficiently consistent. In any case, the age range of the present sample is 25-55 years, which is certainly not an age group where significant deterioration of cognitive functions known as ageing occurs.

general cognitive ability and working memory tests with the influence of age statistically removed (partial Pearson correlation on ranked data; Iman and Conover, 1979). The relationship between experience and the arrow test and the Cattell test are no longer significant and the relationships are weaker, indicating that age was indeed the variable responsible for the relationship.

The following rows of Table 5.5 show correlations with the working memory tests *per se*. Examining the first column, we see that all working memory tests are negatively correlated with age, meaning that higher age is generally associated with worse performance on the working memory tasks. This is again consistent with deterioration of cognitive functions with increased age. There is, however, one exception: the arrow flanker test, where higher scores mean worse performance. If this working memory function deteriorates with age, it would be expected to see a positive correlation. The correlation coefficient is not very high, nor is the relationship statistically significant, so we cannot put too much emphasis on it. It is a deviation from a pattern, nevertheless.

Looking at the second and third columns, we can now look at the relationship between experience (in years and days) and working memory functions. First, the three span tasks, the letter span, the Corsi task and the complex span task, all show negative, statistically non-significant relationships. This pattern is similar to the pattern seen in the first column, which shows correlations with age. In other words, with increased experience, the performance on the span tasks decreases (or, since the relationship is not significant, it is more appropriate to state that there is no relation between the performance on the tasks and interpreting experience). Once the relationship with age is removed (Table 5.6), the relationship is essentially reduced to zero. This particular finding is very interesting and important. In Chapter 3, we reviewed available literature in the area of working memory and interpreting. As discussed, most of the empirical studies set out to test a hypothesis of interpreters' superior working memory in comparison to non-interpreters. While this is not the primary question in this dissertation, the present data can be evaluated from the perspective of the superiority hypothesis as well. Padilla et al. (1995) formulated the hypothesis in terms of enlargement of working memory capacity through training and practice, predicting that experienced interpreters will demonstrate larger working memory capacity than interpreting students and non-interpreters. Our data do not support this view. If working memory capacity were to be enlarged with practice, we would expect to see a positive correlation between experience and the span tasks: more experience associated with better performance on the working memory tests. It is important to point out that neither do our data disqualify the hypothesis. First, no control group is included in the present study, and it may very well be possible that those who become interpreters enter the training with a larger-than-normal working memory capacity and retain it. In such a case, interpreters would still demonstrate superior working memory to non-interpreters, a comparison we cannot make. Secondly, working memory capacity may be enlarged through training, as Padilla et al. (1995) suggested, but this change would be constrained in time (perhaps take place rapidly during training), after which working memory capacity would level off and stay constant. A third possibility is that there is a general decline of working memory capacity in normal population, but that due to interpreters' intensive use of working memory, the decline is slower than in non-interpreters, and demonstrated here by the non-significant negative trends. The superiority hypothesis is empirically contentious, and the indirect evidence presented here is but one small piece in the puzzle.

The last five rows in columns two and three show the relationship between experience and the executive functions of working memory. The antisaccade task measures the ability to inhibit automatic response. The pattern in the data is similar for age and both measures of experience: negative relationships, suggesting decline of the function with increased age/experience. For age, this correlation is statistically significant, for experience measured in years, it approaches significance. Once the interfering influence of age is removed, there is no relationship to experience. Next, the arrow flanker shows a negative relationship, as it does with age, but for both measures of experience, the relationship is both stronger and statistically significant. Arrow flanker is the only working memory test in the present battery where lower scores mean better performance, and the correlations suggest that this particular function is not strongly related to age, but is related to interpreting experience. After the influence of age is removed, the relationship remains strong and significant, which indicates that more extensive interpreting experience is associated with better performance on the task, and hence with better ability to resist interference from task non-relevant stimuli.

The next task is the Barrouillet task, a measure of the ability to switch attention between two tasks during dual tasking. This is the first task where patterns in the relationship with experience do not correspond to patterns in relationship with age. Both experience measures show a very small relationship with the executive function, but slightly positive. The same holds for the last two tasks, the 2-back task and the number-letter task, where there is no pattern: each experience measure shows a different trend; correlation coefficients are very small. One exception is the relationship between experience measured in days and the number-letter task, which shows a small positive correlation which may improve after removing the effect of age. This is indeed the case. The number-letter task and the Barrouillet task both show moderate relationship with experience after controlling for age. The two tasks are operational measures of the ability to switch between two independent tasks and the ability to switch attention during dual tasking respectively, and the data thus indicate that more experienced interpreters are better at controlling their attention and allocating it according to task needs.

The English knowledge test does not show any relationship with working memory functions. With the exception of the Corsi task, all the correlation coefficients are weak, indicating no relationship. The arrow test correlates negatively with all working memory tests and the Cattell test, indicating that faster individuals (lower score on the arrow test) perform better on working memory tests. The relationship is generally stronger with span tasks than with the executive tasks. Similarly, the Cattell task, a test of general cognitive ability, similar in format to abstract intelligence tests, seems to be more related to span tasks (letter span, Corsi, complex span) than to executive control tasks.

This concludes the first examination of the working memory functional structure in a sample of 28 interpreters. The main patterns in the data indicate that working memory functions are fairly independent, as they generally do not correlate with each other. Secondly, age has emerged as a variable which strongly correlates with interpreting experience and acts as an interfering variable, which will need to be controlled for. Thirdly, age and experience are mostly negatively related to working memory functions, indicating that cognitive performance declines with age. The negative correlations with experience are largely reduced once age is controlled for. Only one working memory function has so far emerged as being related to interpreting experience: the ability to resist interference from irrelevant stimuli, as measured by the arrow flanker test. Two other tasks, Barrouillet and number-letter, show indications of being related to interpreting. Finally, the pattern of relationships discussed here goes against the hypothesis of continuous improvement of working memory capacity in interpreters due to extensive practice.

5.3. Structure of simultaneous interpreting

Let us now turn to an examination of participants' performance on the simultaneous interpreting tasks. A total of nine measures of interpreting were taken (Table 5.3). The design of the study assumed that the individual measures will reflect several functionally different types of processes: local processes (lexical, semantic, syntactic processing) and global processes (vocabulary use, planning and speed). Table 5.7 shows the correlation matrix of the simultaneous interpreting measures. Several emerging patterns are visible. The measures of local lexical, semantic and syntactic processing, operationalised as the number of correctly interpreted figures, correctly resolved double negatives and preserved subject-verb agreements in complex syntactic structures, seem to be related. The relationships are positive, the double negatives resolution is significantly related to the other two measures. Interpreters who do well on one of these measures tend to do well on the other measures too. The relationships are not very strong, but they are strong enough to ask whether these measures tap independent processes. This is further supported by the strong positive correlations with the measure companies-average. This measure of interpreting accuracy is taken on lists of figures and noun phrases (lists of human rights and business sectors), which require mostly transcoding (Seleskovitch, 1968/1978), a form of translation where concepts are replaced by their linguistic equivalents in another language, with little processing beyond retrieval from a lexicon. The variables syntax and negatives, on the other hand, require more elaborate forms of processing involving formal grammatical structure and semantics at a sentence level. This positive relationship may thus indicate a more global interpreting ability.

Syntax, figures, negatives or companies-average, on the other hand, do not seem to be significantly related to vocabulary, although the relationships are generally positive and of the same magnitude.

Table 5.7. Correlation matrix (Spearman) of simultaneous interpreting tasks

			Vocabulary		Companies		EVS	
	Figures	Negatives	Ratio	Unique	Difference	Average	Median	SD
Syntax	.248	.338	.440	.337	.078	.626*	-.568*	.-.230
Figures		.486*	-.188	.347	-.231	.680*	-.501*	-.439*
Negatives			.359	.341	-.074	.537*	-.498*	-.337
Vocabulary Ratio				.287	.214	-.019	-.243	-.215
Vocabulary Unique					-.025	.291	-.013	-.100
Companies Difference -						-.363	.154	.260
Companies Average -							-.567*	-.378
Median EVS								.651*

* p< .05

Another very interesting pattern emerges in the relationship with median ear-voice span. All three measures of local processing are negatively and significantly related to EVS. This means that those interpreters who kept shorter EVS were more successful in interpreting the measured items correctly. This is not a new finding. The literature on ear-voice span discusses the issue of striking the right balance between waiting in order to listen to more text, resolve ambiguities etc. on the one hand, and interpreting as soon as one can in order not to overload the limited processing resources (de Groot, 1997, Setton, 1999, Goldman-Eisler, 1972). Lee (2002) tested empirically the relationship between accuracy and EVS, and reported higher level of accuracy in segments with EVS shorter than two seconds and reduced accuracy in segments with EVS longer than four seconds. This finding is supported by the present data, with shorter EVS associated with higher level of accuracy.

What is perhaps more interesting is the relationship between companies-average and median EVS. The relationship is again significant and negative, similar to the relationship with syntactic, semantic and lexical processing. The difference here is that EVS and companies-average measurements are independent, as they were each made in a different text, while the three measures of local processing and EVS were measured on the same items. The fact that EVS is related in the same fashion to accuracy in another text strongly suggests that EVS measures an underlying process which transcends the immediate context of a given interpreting task and that it could potentially be a more stable characteristic of interpreters' behaviour. This result is very interesting and will be subjected to scrutiny in further analyses.

Next, we performed a series of t-tests to examine differences due to sex and due to mother tongue. There were no differences between males and females on any of the interpreting measures. As for mother tongue, there was a significant difference between Czech and Dutch interpreters on the average accuracy in the companies texts[24] and a marginally significant difference on the difference in accuracy between the companies texts.[25] Dutch interpreters achieved higher accuracy in the interpretation of the companies texts, and were marginally significantly less affected by speed, than Czech interpreters. There were no other differences associated with the interpreters' mother tongue. Other participant characteristics (age, experience, general cognitive ability and knowledge of English) and their relationship to simultaneous interpreting performance are shown in Table 5.8.

First, we note that the pattern of correlations with age and experience differs from the pattern seen in the working memory tests. There, age was the more strongly related variable, with experience correlating more weakly and less often significantly. In the simultaneous interpreting tasks, on the other hand, experience is the variable which correlates more strongly, and age usually produced correlations in the same direction, but weaker. This suggests that relationship with age will be driven primarily by experience, and once experience is removed, age may not show any relationship at all. This is indeed the case, as shown in Table 5.9. After the experience in years has been statistically removed, age does not

[24] M_{Czech}=49.47, SD_{Czech}=9.45, M_{Dutch}=57.93, SD_{Dutch}=8.04, $t(24) = -2.10$, p=.047
[25] M_{Czech}=10.11, SD_{Czech}=6.39, M_{Dutch}=5.29, SD_{Dutch}=5.62, $t(24) = 1.76$, p=.09

correlate significantly with any SI measure. It is also interesting that correlations with Cattell, the test of general cognitive ability, is often negative, which would suggest that interpreters with higher general cognitive ability do worse on interpreting tasks than interpreters with lower general cognitive ability. That is highly counterintuitive, and an obvious possibility is that this is again a hidden effect of age.

Table 5.8 Correlations (Spearman) between SI measures and personal characteristics

	Age	Experience years	Experience days	English test	Arrow test	Cattell
Syntax	.397*	.438*	.372	.159	.305	-.234
Figures	.082	.271	.237	.164	-.032	-.143
Negatives	.396*	.587*	.624*	.240	.031	-.204
Vocabulary Ratio	.349	.400	.256	.279	-.005	-.332
Vocabulary Unique	.344	.427	.269	-.025	.305	-.286
Companies Difference	-.376	-.324	-.385	-.391*	-.074	.185
Companies Average	.385	.524*	.510*	.336	.189	-.050
Median EVS	-.216	-.360	-.434*	-.382*	.101	.184
EVS SD	-.196	-.414*	-.412*	-.524*	.207	.153

* p< .05

Table 5.9 Partial correlations between SI measures and age controlling for experience in years, and between SI measures and experience controlling for age (Pearson on ranked data)

				Vocabulary		Companies		EVS	
	Syntax	Figures	Negatives	Ratio	Unique	Difference	Average	Median	SD
Controlling for experience (years)									
Age	.063	-.293	-.206	-.004	-.041	-.149	-.152	.161	.294
Controlling for age									
Experience years	.213	.388**	.505**	.252	.304	-.068	.422**	-.332*	-.461**
Experience days	.147	.288	.528**	-.022	-.022	-.208	.390*	-.404**	-.391**

** p<.05, * .05 < p < .10

Next, both measures of experience are fairly strongly related to a number of measures of simultaneous interpreting, both indicating the same – and predictable – direction. More experienced interpreters thus tend to achieve higher score on measures of syntactic, lexical and semantic processing, produce more unique words and a text with higher vocabulary ratio, they tend to have shorter EVS, less variability in their EVS and a smaller difference between texts delivered at different speeds, indicating better ability to cope with speed. However, once age is controlled for (Table 5.9), the pattern changes for several SI measures. Most interestingly, two measures of local processing are affected in opposite ways. Syntax is no longer related to experience measured in years, while figures become significantly related to experience in years. A third measure with an unusual relationship to age is companies-difference, which seems more strongly related to age than to experience. A closer examination of the three SI variables indicates that there may be some age-related changes in interpreting performance. This effect is most clearly visible on a scatterplot of age and companies-difference shown in Figure 5.2. Similar, although less pronounced, relationships exist between age and each of syntax and figures, suggesting non-linear trends in the data, which are not accurately captured by the linear correlation statistics.

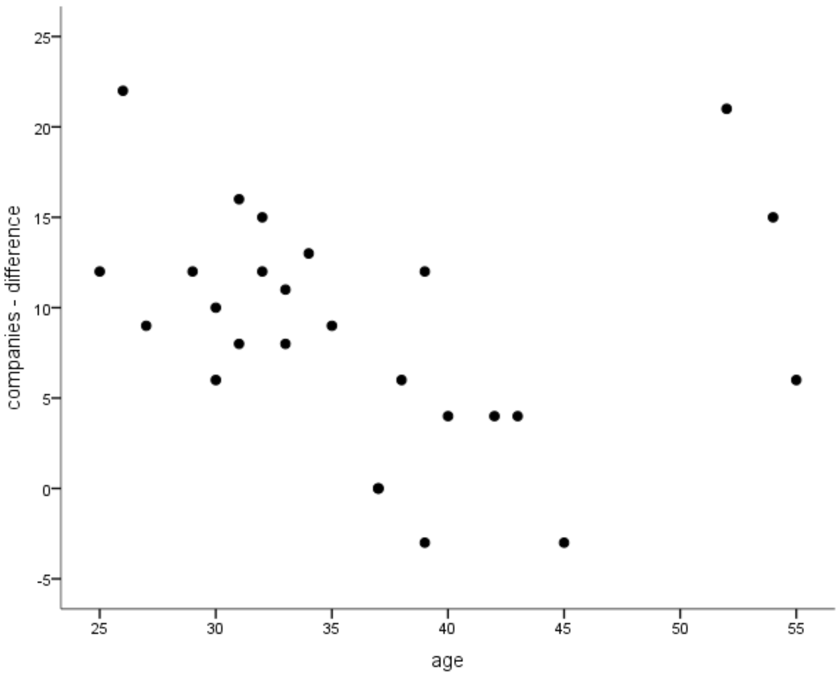

Figure 5.2 Relationship between age and the average number of correctly interpreted items in the companies texts

To examine the data further, we will consider figures, as numbers are a major component of the score on the companies texts as well, and are relatively easy to assess for various effects. One possible explanation behind the relationship pattern with experience (with age included and excluded) is a change of strategy by more experienced interpreters. This directly concerns the scoring method employed in the present study. Only figures which were correctly interpreted in their entirety scored a point. Figures which were interpreted only approximately correctly were considered incorrect and awarded zero points. For example, if the source text contained the figure *46.5*, interpretations such *46*, *more than 40* and *46.3* were all scored as incorrect. Common interpreting norms, however, allow approximations, and it is possible that interpreters, especially those with more experience, employed this technique in their interpreting. Assessing any form other than the exact equivalent is very difficult, as it cannot be determined with certainty whether it is the result of a strategy employed by the interpreter, or whether it is a genuine error. It can be argued, however, that both *46* and *more than 40* in the above example are interpretations based on a strategic decision by the interpreter. The third example *46.3*, on the other hand, can be considered to be a genuine error, as it contains elements which are not motivated by the source text. If experienced interpreters employ the strategy of rounding numbers or providing other types of approximations[26] when they judge the exact figure to be unimportant for achieving the goal of message transfer, we would expect their performance to mostly contain correct and approximate renditions[27]. By the same logic, we would expect no or very few errors. To test this prediction, we have run a supplementary analysis on figures in the Amnesty text. A total number of errors were counted per interpreter. Correctly interpreted figures, approximations of any kind and omissions were not considered; only genuine departures from the source text were included. The mean number of errors was M=2.2, SD=2.1 (out of 24 figures), with a minimum of zero errors and a maximum of eight errors. The number of erroneously interpreted figures was correlated with experience in years and in days. The Spearman correlation coefficient was not significant for experience measured either in years or days[28], and Figure 5.3 shows the scatterplot of the data. Both the numerical and the graphical information indicate that experienced interpreters do not make fewer errors than less experienced interpreters. As one more complementary analysis, we checked the number of approximations in the interpretations to see whether there is a legitimate case for substantial improvement of the score by including the approximate figures. Only 18 cases were found, and if the approximations were included, 14 interpreters would increase their score by one point, 2 interpreters by two points, and 12 interpreters would keep the current score. The approximations affected only 18 out of a total of 672 figures, or 2.7% of the data, and hence do not seem to be an important phenomenon, which would affect the results in Table 5.3 to any meaningful degree. The lack of correlation between correctly interpreted figures (using the strict scoring procedure as defined in the method section) and interpreting experience, a similar lack of correlation between the number of errors and interpreting experience, and

[26] For example, *approximately half* is an acceptable approximation to *46%*, *many millions* for *35 million*, etc.
[27] There may also be strategic omissions, but in the absence of any interpretation at all, it is very difficult to judge whether an omission was the result of a strategy or a genuine error.
[28] $r_{s(years)}(28)=.145$, $p=.463$, $r_{s(days)}(28)=.212$, $p=.278$

generally low number of approximations all support the same conclusion: interpreting figures does not seem to be related to interpreting experience.

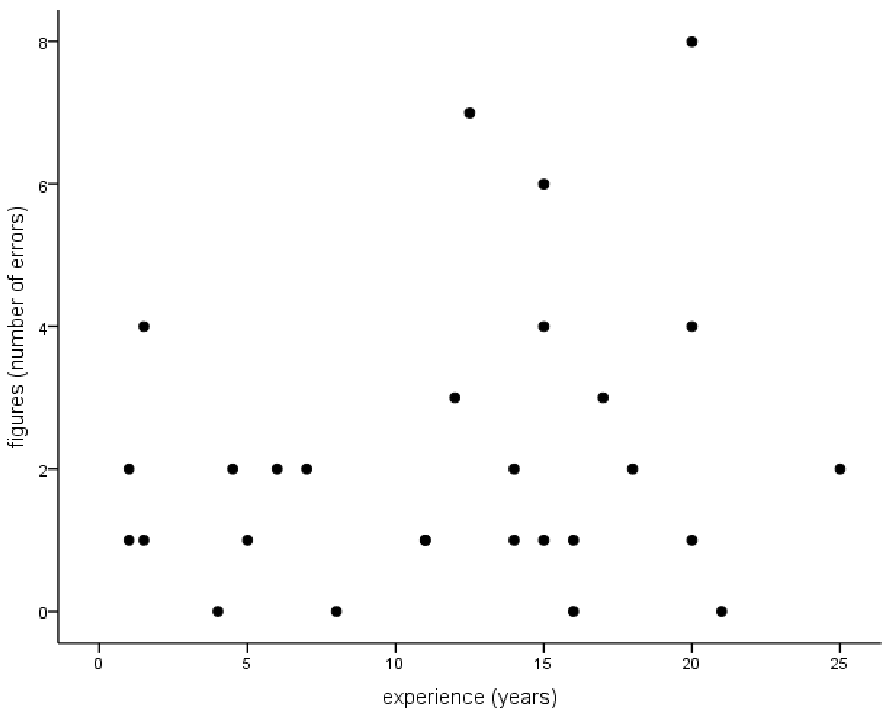

Figure 5.3 Relationship between experience (in years) and the number of incorrectly interpreted numbers

It is difficult to assess precisely what produced this pattern of relationships, and what is the exact interaction between age and experience in interpreting. The unusual values in Figure 5.2 may be simple outliers, or the age-related trend may be real. Our sample is very small and does not cover the whole age range: more participants aged 45-55 would have to be tested to allow a better evaluation of the pattern. We can only conclude that there seem to be age-related patterns in interpreters' performance, which should be studied in more detail in future research. Note, however, that statistical inclusion or exclusion of age affected primarily the relationship between SI tasks and experience in years, while the relationship with experience in days was affected to a lesser degree. Since experience in days seems to be somewhat less sensitive to effects of age, further analyses will use the experience in days measure.

Let us continue our examination of SI structure. In Table 5.7, we saw a correlation between EVS and accuracy measured in different texts and we suggested that EVS and the ability to deal with speed may be a characteristic of interpreters' performance, which goes beyond a specific text. The relationship between EVS and experience shown in Table 5.8 may offer one explanation: with increased experience, interpreters are able to work with shorter EVS, which allows them to keep up with the speaker and produce more accurate output[29]. This explanation could be further supported by the negative correlation between English knowledge and EVS and English knowledge and companies-difference. Better knowledge of English is associated with shorter EVS and smaller difference in interpreting texts under different speed conditions. That could suggest easier retrieval, and hence faster processing time and shorter EVS.

The observed negative correlation with variability in EVS (EVS SD) is a finding contradicting the initial conceptual design of this study. EVS was included as an operational measure of planning, with the assumption that interpreters will vary their ear-voice span as a function of text demands. Relevant literature suggests that experienced interpreters use EVS as a strategic tool (Gile, 1997) and that they work on a more global level, with longer EVS (Moser-Mercer, 1997a – Beyond curiosity), with EVS. This led us to hypothesise that more experienced interpreters would engage in more planning and consequently demonstrate longer EVS and likewise more variability in function of the currently processed segment. Our data do not support this interpretation. Instead, EVS was found to be shorter and more constant in interpreters with more experience. A possible explanation in theoretical terms is offered by Moser-Mercer (1997b – expert novice paradigm) in terms of better knowledge organization in experienced interpreters resulting in faster access and retrieval. On an empirical level, de Groot (1997) analysed previous research on EVS, which yielded inconsistent results. Interpreters sometimes showed preference for a constant EVS over accuracy in adverse conditions, such as noise added to input (Gerver, 1974, cited in de Groot, 1997). But when density and input rate were manipulated, EVS was longer. Our study employed materials free of noise (to the limit afforded by available technology), with very high information density and input rate in the range of 120-130wpm. It is difficult to explain the locus of the difference between the present findings and previous research, as there are a number of important differences between the studies; chief among them is the non-experimental inter-subject comparison made on a single text in the present study, and the experimental design comparing intra-subject performance in a number of conditions. Perhaps the best explanation is that the input demands of the present testing materials were such that they encouraged as fast processing as possible in order to keep up with the speaker. In this sense, the individual EVS may reflect the limits of speed with which an interpreter is able to interpret, and it is possible,

[29] It should be kept in mind, however, that “accuracy” as measured in this research project, is fairly removed from “accuracy” as ordinarily understood among professional interpreters.

even likely, that our participants would have produced different patterns of EVS on input speeches with different characteristics.

A third pattern, and perhaps the most interesting one, concerns figures. We saw in Table 5.7 that the three measures of local processing, syntax, negatives and figures, were related to each other, which suggests they do not tap separate processes. Here, however, we see that interpreters' scores on the number of figures correctly interpreted is related differently to the other variables. First of all, and most importantly, it does not seem to be very strongly related to experience, unlike the other two measures, syntax and negatives. In fact, figures have fairly low correlations with all other variables, and no immediate relationship is apparent here. Figures are related to EVS (as shown in Table 5.7), which suggests they are sensitive to speed, but interpreters do not seem to become better at coping with figures with more experience. Let us examine in more detail the relationship with ear-voice span. A complementary analysis, shown in Table 5.10 and Figure 5.4, was carried out on the relationship between the accuracy of individual figure items (i.e. each of ten sentences containing a total of 24 figures) and ear-voice span measured at the beginning of each sentence. A total of 280 sentences were measured (10 sentences by each of 28 interpreters). A total of 273 sentences were interpreted, seven sentences were omitted in their entirety and do not have an associated EVS measurement. Of the 273 sentences, 52 had no correctly interpreted figures (following the strict scoring procedure, this category includes omissions, approximations and true errors), 83 had one correctly interpreted figure, 95 sentences had two figures and 43 sentences had three correctly interpreted figures.

Table 5.10 Correctly interpreted figures (means, standard deviations and 95% confidence limits on the mean)

Number of correct figures	N	M	SD	95% confidence interval for mean	
				Lower limit	Upper limit
0	52	4.4	2.5	3.8	5.2
1	83	4.1	2.0	3.7	4.6
2	95	3.1	1.6	2.8	3.5
3	43	2.6	1.9	2.0	3.1
Total	273	3.6	2.1	3.3	3.8

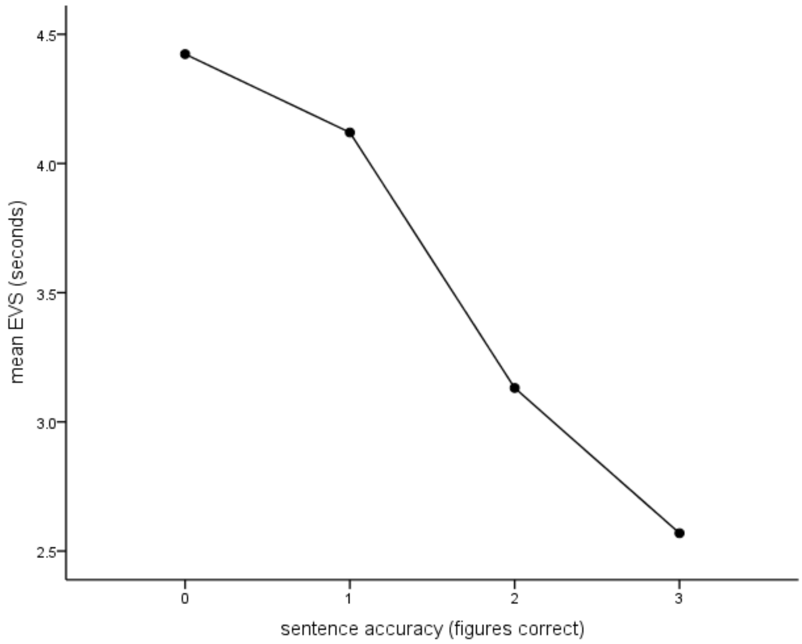

Figure 5.4 . Relationship between the number of correctly interpreted figures in a sentence and EVS

The mean EVS on each category of sentences according to the accuracy of interpreted figures is shown in column 3 and Figure 5.4. The 95% confidence interval for the mean suggests two homogenous groups of sentences – those with zero or one correctly interpreted figure, and those with two or three correctly interpreted figures, as there is no overlap of the confidence intervals between these two groups. This indicates a significant difference in EVS, with EVS being shorter on sentences with two or three correctly interpreted figures, and longer on sentences with none or one correctly interpreted figures, with the line being drawn – in the present data – at ear-voice span of approximately 3.5s. We can therefore conclude that full and accurate interpretation of figures is more strongly related to EVS than to interpreting experience.

To summarise, in this section we have examined the structure of the simultaneous interpreting task. We have detected several patterns in the data. First, there is generally a relationship between the measures of local processing (syntax, negatives, figures), with figures possibly falling into a separate subcategory due to the slightly different pattern of relationships to other variables than that shown by syntax and negatives. Secondly, performance on the companies texts and general EVS are related to measures of local processing, indicating that better adjustment to input speed results in better performance on other variables. Also, the relationship between companies-average and EVS indicates that these characteristics may be a more constant feature of an interpreter’s performance. Vocabulary measures, on the other hand, were largely uncorrelated with the other measures and seem to form a separate component of the interpreting task. Finally, measures of the interpreting tasks are more strongly related to interpreting experience than to age or general cognitive ability, which emphasises their acquired nature developed through practice, although several patterns are present in the data which indicate that a complex interplay between age and experience may be taking place.

5.4. Working memory and simultaneous interpreting

Having considered the structure of working memory and simultaneous interpreting in the present sample of 28 interpreters, let us now proceed to examination of the relationship between performance on simultaneous interpreting tasks and interpreters' working memory. Table 5.11 shows the correlation matrix between the two constructs. In contrast to the separate analysis of each construct, we first note a lack of consistent patterns in the correlations. This needs to serve as a warning that the data should be examined with maximum caution, as the chance of the relationships being spurious is higher.

Table 5.11 Correlation matrix (Spearman) of working memory and simultaneous interpreting tasks

	Letter	Corsi	Complex span	Antisaccade	Arrow flanker	Barrouillet	2-back	Number-letter
Syntax	.193	-.120	-.190	-.170	-.301	.055	-.039	-.124
Figures	.337*	-.184	.183	-.062	-.092	.275	.522**	-.409**
Negatives	.026	-.108	-.114	.088	-.304	.372*	.093	-.024
Vocabulary ratio	-.301	-.103	-.370	-.091	-.142	.193	-.648**	.201
Vocabulary unique	-.180	-.368	-.284	-.173	.013	.334	-.109	.115
Companies difference	.106	-.068	.006	.167	.137	.543**	-.126	.205
Companies average	.372*	-.223	.061	.003	-.423**	.099	.312	-.216
Median EVS	-.013	-.031	.071	.071	.092	-.227	-.266	.415**
SD EVS	-.068	-.101	-.009	.141	.156	-.055	-.152	.116

**p<.05, * .05 < p < .10

First, let us consider the first three columns of data, which contain correlations between the three measures of memory capacity (letter span, Corsi task, complex span) and measures of simultaneous interpreting. Only letter span, a measure of verbal short-term memory without a processing component, shows a relationship with simultaneous interpreting, but the relationship only approaches significance. The Corsi task, a measure of visuospatial memory, and the complex span task, a variant of complex working memory tasks of the reading span type, do not seem to be related to any of the measures of interpreting selected for this study. The letter span task, too, is related only weakly and statistically marginally significantly to two measures of simultaneous interpreting: the number of correctly interpreted figures and the average number of correctly interpreted items in the companies texts. In both cases, the task consists of horizontal translation (Seleskovitch, 1968/1978), a process further enhanced in the present data by the selected scoring method, which only accepted exact matches as correct interpretations. The verbatim memory demands of the letter span task provide a plausible explanation for the observed relationship, and also an explanation of why no such relationship is present between, for example, letter span and negatives, where the interpreting task requires analytical processing, rather than simple storage and transcoding.

A much more interesting picture is offered by the central executive tasks. Of the five tasks tested in this study, four show relationships with simultaneous interpreting, two of them on two measures each. First, figures are correlated with two measures of central executive functions, the 2-back task, a measure of updating, and the number-letter task, a measure of the ability to switch between tasks. In both cases, better performance on the central executive task is associated with more correctly interpreted figures. Interpreters who correctly interpreted more figures were also better able to update their memory and were faster at switching from one task to another. The 2-back task is also negatively related to vocabulary ratio: interpreters who were better at updating their memory tended to have a lower type/token ratio. This relationship is difficult to explain in conceptual terms. Those who updated their memory more efficiently tended to choose less variable vocabulary. This could possibly point in the direction of greater demands on processing speed. Such an explanation is, however, based on knowledge of the mechanics of the 2-back task and interpreting. It is much more difficult to support it empirically; neither the 2-back task, nor the vocabulary ratio score are involved in relationship patterns with other variables, and there is therefore no converging evidence to support the explanation offered. This relationship is perhaps best left with a question mark for further research.

A similarly puzzling relationship has been found between the arrow flanker task (a measure of ability to resist interference from task-irrelevant stimuli) and the average number of correctly interpreted items in the companies texts. The arrow flanker task was correlated with interpreting experience, independently of age, and a correlation with an interpreting variable would further support the relationship. It is rather difficult, however, to find a meaningful explanation for the negative correlation with the companies - average measure. Our best estimate of a possible reason is that a successful performance on the companies texts – texts consisting largely of strings of unrelated items – might require very tightly focused attention on the incoming string of figures and concepts.

The last two relationships present in our data are easier to explain. The difference between the two companies texts attempted to measure the extent to which interpreters are affected by speed; the Barrouillet task measures essentially the same thing – how one's performance (on a memory task) is affected when one is prevented from switching attention to the task by being engaged in another task. In both cases, the relative increase in attention demand is achieved by increasing speed. This explanation for the relationship is theoretically plausible, and tentatively ties in with the explanation offered for the relationship between companies-average and arrow flanker. More specifically, the demands of the Barrouillet task are such that if attention is captured by the processing task, it cannot be switched to the memory task, and the memory task suffers. In the interpreting task too, there are (at least) two tasks: processing of the incoming message and production of the target output. To succeed under the unfavourable conditions of high input speed, the interpreter either has to switch attention very efficiently between input and output, or leave production to autopilot and focus on input. Automated production in the specific context of the companies texts is in our opinion possible, as the texts contained mostly unrelated strings of items, or items embedded in rather formal context without substantial information value. If the ability to control attention and selectively direct it to input were a skill developed with increased interpreting experience, this would explain the relationship between interpreting experience and both the arrow flanker task and the Barrouillet task (which emerged after age was controlled for, see Table 5.6). It needs to be pointed out that the explanation is proposed on the basis of empirical data collected in this study, and without further testing it is no more than a way to rationalise about the observed pattern.

The last relationship in Table 5.11 concerns the number-letter task, a measure of efficiency with which attention is shifted from one task to another, and the ear-voice span. This relationship too, has an intuitive theoretical explanation. Interpreters who are better able to switch attention also keep shorter EVS. Finding this relationship is exciting: in the structure of simultaneous interpreting tasks, EVS was a variable related to a number of measures of SI, suggesting that it may be an explanatory variable behind a number of phenomena. This was further supported by the relationship between EVS and accuracy of figures, where shorter EVS was associated with higher accuracy. The relationship between EVS and the efficiency of shifting from one task to another might provide a further explanation of the processes behind the association of EVS and accuracy of figures.

To provide a more complete picture, the same analysis (correlations between working memory and simultaneous interpreting measures) was run while controlling for the effects of age, which turned out to be a nuisance variable in the analysis of the structure of simultaneous interpreting. The essential picture does not change: relationships identified here were preserved, those marginally significant became significant, and a few more marginal relationships appeared. The full matrix is not reproduced here as it does not add much new information. One result, however, has changed substantially: the relationship between arrow flanker and companies-average was weaker and no longer significant (although similar in broad terms in both direction and magnitude). Since arrow flanker was related to experience, but not age, it is not entirely clear what produced the result. It does, nevertheless, stress again the importance of more research into the effects of age. Importantly, with effects of age removed, arrow flanker was no longer related to a single measure of SI. Its relationship with experience suggests that its effect should be reflected in SI performance, but perhaps in another measure. A supplementary linear regression on ranked data with companies-average as a criterion, and age and arrow flanker as predictors, shows that there is a significant interaction between age and arrow flanker[30], indicating that age is a moderating variable. Since age is strongly related to experience, as is arrow flanker, the interaction may be an indication of the developing ability to resist interference from irrelevant stimuli.

The present results immediately raise two other questions. First, it cannot be expected that any of the simultaneous interpreting measures would be explained by a single variable, either from the interpreting domain, such as experience or EVS, or the working memory domain. Interpreting is a far too complex activity to rely on a simple one-to-one relationship. Secondly, several associations appeared in the data between experience and both simultaneous interpreting and working memory. It would be very interesting to probe these further in order to understand the interaction between basic cognition and interpreting experience, between nature and nurture. The present data set is unfortunately too small to allow for a meaningful analysis of multivariate relationships. For the sake of exploration, we will present a few carefully selected results of analyses, which examine the relationship between more than two variables. The results should be viewed as extended exploration only, and certainly not as confirmatory.

[30] $F(3,21) = 5.135$, $p = .008$, Adj $R^2 = .341$, $\beta_{\text{arrow flanker}} = -.408$, $p = .031$, $\beta_{\text{age}} = .148$, $p = .420$, $\beta_{\text{arrow flanker*age}} = .452$, $p = .015$

Table 5.12 shows four regression models with two predictors each. All models are hierarchical to allow the assessment of the contribution of individual predictors to explaining the overall variation in the criterion scores. All analyses were run on ranked data (Iman and Conover, 1979) to reduce the influence of extreme values. To maximise use of available data, missing values were excluded pairwise. The first two models test whether a measure of simultaneous interpreting can be predicted from two separate measures with which it was correlated. The second column for each predictor (ΔR^2) indicates the amount of variance explained by each predictor, and the last column shows the total variation in the criterion scores explained by the two predictors together. The adjusted R^2 was chosen for reporting, as it provides a more conservative estimate appropriate for small samples (Howell, 2010). The companies average score correlated with EVS (Table 5.7) indicating that a higher number of correctly interpreted items is related to shorter time-lag between the speaker and the interpreter. The same score was also related to the arrow flanker task (Table 5.11), suggesting that performance on this specific task is also related to better ability to resist interference from task irrelevant stimuli. The two predictors, EVS and arrow flanker, taken together explain a total of .390, or 39%, of variation in the companies-average scores, which is substantially more than was explained by either of the two predictors separately. The independent contribution of the two predictors is significant (shown by the significant standardised regression coefficients β). Since arrow flanker was found to be related to interpreting experience (Table 5.6), so that better performance on the task is related to longer interpreting experience, which suggests that interpreters, through their interpreting practice, develop their ability to inhibit irrelevant stimuli, which then contributes to better performance on a simultaneous interpreting task.

Table 5.12 Selected two-predictor hierarchical regression models

Criterion (DV)	Predictor 1			Predictor 2			Model
	Name	ΔR^2	β^a	Name	ΔR^2	β	Total adjusted R^2
Companies-average	Arrow flanker	.174*	-.370*	EVS median	.266*	-.518*	.390*
Figures	2-back	.281*	.432*	EVS median	.144*	-.393	.378*
Experience days	Arrow flanker	.359*	-.564*	EVS median	.145*	-.382*	.462*
EVS median	Number-letter	.177*	.410*	Experience days	.179*	-.423*	.303*

[a] β for predictor 1 refers to the two-predictor final model
*$p<.05$

A similar pattern is shown in the second row, where figures were predicted by EVS and the 2-back task, which measures the ability to hold information and update it according to task demands. Again, the two predictors contribute independently, jointly explaining 37.8% of variation in the figures scores. The difference between these two models is that 2-back was not found to be related to other measures of working memory or interpreting, including experience. This suggests that, at least within the set of variables included in this study, the 2-back task measures an interpreter's natural cognitive ability, which contributes to interpreting performance, unaffected by interpreting experience.

The first three models include EVS as a predictor. Additionally, EVS correlates with simultaneous interpreting measures related to lexical, semantic and syntactic processing (although not to vocabulary measures). The multiple association of EVS with other measures suggests EVS could be a common factor in all these relationships, a "supervariable" of a kind. We have noted its relationship to experience, such that EVS is shorter with more experience, and also with a measure of working memory central executive function, the ability to switch between tasks, as measured by the number-letter task. The fourth regression model presented in Table 5.12 tests whether EVS can be predicted from these two variables. Both experience in days and the number-letter task predict EVS independently, jointly explaining about 30% in EVS variation among participants. This indicates that EVS, too, can be explained by a general cognitive component – the ability to switch efficiently from one task to another – and a learned component acquired through practice.

To summarise, in this section we presented the results of a correlational analysis between the two constructs, working memory and simultaneous interpreting. Several relationships were identified between various measures of simultaneous interpreting and several working memory measures, predominantly those related to central executive functions. The results indicate that interpreting performance can be predicted on the basis of WM measures, and that prediction becomes better when several predictors are used. Most of the WM measures that correlated with SI performance are unrelated to experience and suggest therefore that interpreters' cognitive makeup may play a role independently of the mastery of the acquired skill.

Chapter 6. Conclusions

This dissertation set out to explore the relationship between working memory and simultaneous interpreting, motivated by long-standing claims about the existence of such a relationship on the one hand, and paucity of empirical support on the other. Before we discuss the specific research questions which guided the present empirical study, let us consider some methodological aspects of the study.

6.1. Methodological limitations

This is an exploratory study, aiming to map the relationships between working memory and interpreting. In the absence of solid theoretical predictions, the study was designed very broadly and also provides results which need to be interpreted in the broadest terms. It is important to stress that the conclusions are data-driven, i.e. theoretical explanations are offered in order to make sense of the observed phenomena, which, however, does not mean such explanations are not necessarily true. They should only be considered as proposals and independently further tested.

Some patterns observed in the present data suggest that the results may be influenced by the nature of the simultaneous interpreting testing materials. Chief among these is that the interpreting tasks clearly favoured fast and efficient processing over linguistic and stylistic creativity. Substantially different testing conditions may thus give rise to very different results. A similar limitation is related to the simultaneous interpreting variables selected here. They represent only a fraction of interpreting phenomena that could have been selected but the motivation behind the selection was explained in Chapter 4. Even with the variables selected, different ways of embedding them in the interpreting tasks and different methods of scoring may lead to substantially different outcomes. For example, the time lag between the speaker and the interpreter was scored in two different ways – we measured the median EVS, as a measure of the most typical lag each interpreter kept, and the standard deviation of EVS, as a measure of the variability in individual EVS measurements. Both measures were based on the same data, but lead to quite different results.

Additionally, age turned out to be a nuisance variable which seems to have an effect on a variety of phenomena but is very difficult to statistically control for in a small sample. Age was more systematically considered by Signorelli et al. (2011) and we agree with their conclusions that age needs to be watched out for and/or controlled. In practical terms, when recruiting “experienced” interpreters, researchers need to consider age separately. Our data suggest an age divide located between 45 and 50 years of age. This figure is being provided as guidance only for more specific research. We need to stress that we do not associate age with decrease in performance, we merely state that the patterns in the data seem to be different in younger and older interpreters. Also, our sample consisted mostly of people younger than 45 years, there are only a few data points beyond this artificial line, and they are insufficient to specify whether the pattern is consistent or spurious. Equally importantly, there do not seem to be age-related changes in all variables, but until more is known about the interaction between age and interpreting-related variables, researchers need to exercise due caution and be aware of the potential confounding of results.

6.2. Research questions

The study addressed three questions:

4. Is there a relationship between working memory and simultaneous interpreting performed by professional interpreters?
5. Is working memory involved in all aspects of simultaneous interpreting to the same extent, or do different functions of working memory support different processes in simultaneous interpreting?
6. How strong is the relationship between working memory and simultaneous interpreting? Do data support the notion that working memory is a crucial mechanism of simultaneous interpreting performance?

Partial answers were given in the discussion of the results of individual analyses presented in Chapter 5. In our conclusions, we will focus on a more global picture and will bring the results together on a more conceptual level. On the basis of our data, we constructed a network of relationships (Figure 6.1) and will use this “map” for the discussion of our main findings. The map shows five white rectangular boxes, which represent simultaneous interpreting measures and experience, and five shaded ovals, which represent working memory functions. Only selected measures and relationships are shown, rather than an exhaustive depiction of all variables, as this enhances clarity and allows for a better understanding of the main patterns seen in our data.

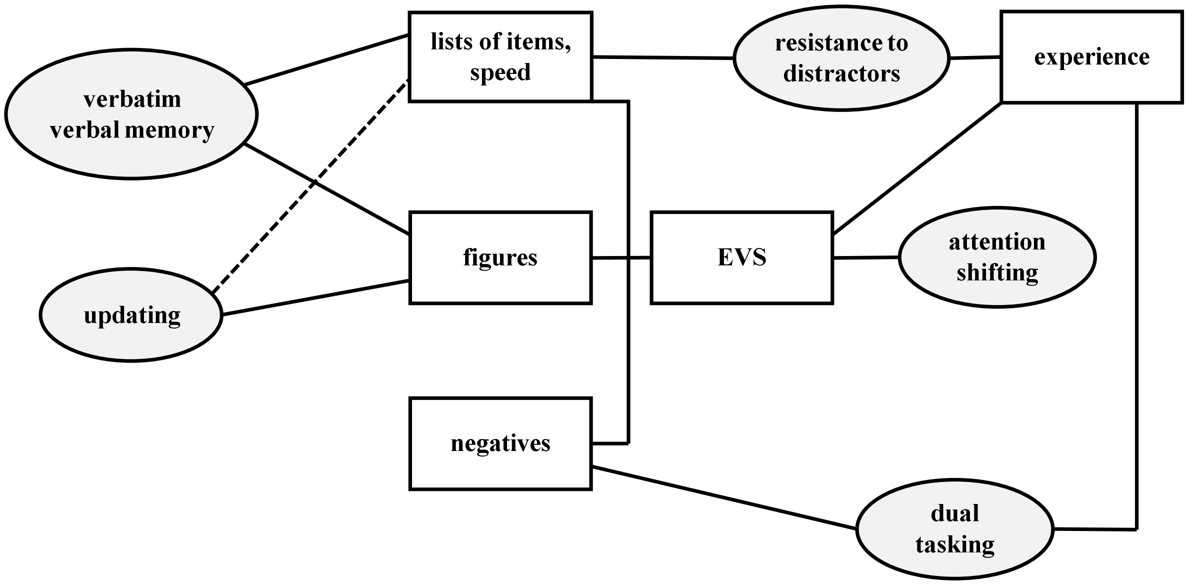

Figure 6.1 The map of working memory involvement in simultaneous interpreting

Several working memory functions (shaded ovals) are shown to be related to several measures of simultaneous interpreting. All the relationships shown in Figure 6.1 are positive – better working memory means better interpreting, which empirically supports the expectations based on theoretical literature. In answer to our first question we therefore conclude that there is indeed a relationship between working memory and simultaneous interpreting performed by professional interpreters.

The network of relationships is fairly complex, with some working memory functions being related to more than one measure of simultaneous interpreting. Consider, for example, the interpreting of figures. Interpretation of figures will be more successful if the interpreter a) has a good verbatim verbal memory to remember the figures while they are being processed, b) is able to quickly update information held in memory in response to task demands, c) keeps short ear-voice span, for which she needs d) the ability to efficiently switch attention between several tasks and e) interpreting experience, which probably reflects a host of other skills and abilities. Similarly complex relationships hold for the other measures of simultaneous interpreting where an association with working memory was found. The pattern of relationships between the individual working memory functions and measures of simultaneous processing can best be described as many-to-many. Some functions and aspects are paired with one or more other functions and aspects, and some did not show any associations at all (those are not included in Figure 6.1). Therefore in response to our second question we conclude that working memory is not involved in simultaneous interpreting as one entity, but that different working memory functions are related to different types of interpreting processes in a fairly complex network of interactions.

The most interesting aspect of the relationship between working memory and simultaneous interpreting concerns the fairly clear division between the *memory* components of working memory and the central executive functions. Of the memory tasks, only the test of verbatim verbal memory (memory for language, either written or spoken) was related to interpreting. The two measures to which it was related are figures and lists of items presented under different speeds. Both figures and enumerations are fairly context-independent, they cannot be easily inferred or reconstructed if missed, and remembering them in their verbatim form may be an effective way of storing them while they are being processed.

The remaining four working memory functions in Figure 6.1 represent the central executive. Updating is a measure of the ability to hold information relevant for the task, and as soon as the task is completed, the interpreter flushes the memory and updates it with information relevant for the next task. There is a small memory component, but the information only needs to be maintained for a few seconds. The main focus of this function is on the ability to effectively update the information, including removing traces of previous information so it does not interfere with the current task. Next, resistance to distractors is the ability to focus attention and ignore irrelevant stimuli. The last two functions concern two aspects of attention in a multi-tasking setting: sharing attention between two tasks (dual tasking) and switching attention from one task to another (shifting).

All four central executive functions depicted in Figure 6.1 are closely related to attention and coordination, and their involvement in interpreting on a general level makes intuitive sense. Simultaneous interpreters are constantly juggling an input stream of speech which needs to be processed and reproduced in another stream of speech. Control of attention, the ability to focus attention on where it is needed, was proposed as a crucial component of interpreting (Cowan, 2000/1, Moser-Mercer, 2005). Liu et al. (2004) observed in their data that more experienced interpreters were better at not missing critical segments. They suggested the ability to switch attention as an explanation, although this proposal was not directly tested. Their findings are broadly in line with those presented in this study and to use Cowan's (2005:viii) terms, our data support a relationship with working memory *capability* as opposed to working memory *capacity*. We would therefore like to propose that the present data add empirical support to the suggestions about the role of attentional control in interpreting. An additional question is whether our support extends to the idea of attentional control being crucial for interpreting. The relationships with central executive functions found in our data are all respectably strong. They are not many, however. One reason may be methodological: the selection of interpreting measures in this study may not have been the best candidates for processes involving working memory, although the selection was made carefully on the basis of available literature. Another possible explanation is that working memory is primarily involved in background processes which are difficult to measure in the product. This view would be supported by the relationship between attention switching as a predictor of ear-voice span, which in turn is related to a whole range of phenomena in the final product. While both are a plausible possibility, with respect to our third question, we conclude that the present data are not sufficient for the attentional control as the central executive component of working memory, to be the most important underlying component of simultaneous interpreting. On the other hand, we are confident to conclude that the storage (memory) functions of working memory are not the key element underlying simultaneous interpreting.

Let us now consider more general implications of the present findings. Interestingly, and importantly, the data indicate that the relationship between simultaneous interpreting and working memory has two distinct paths: one related to experience, and the other related to innate cognitive ability. Figure 6.1 shows that the ability to resist distractors and the ability to switch attention in dual tasking conditions are related to interpreting experience. Both functions could therefore develop with the skill. The other three central executive functions do not show any association with interpreting experience, and we therefore conclude that they reflect innate cognitive abilities. However, if working memory were directly related to simultaneous interpreting, so that it developed with more practice, we would expect to see evidence of the particular function in relation to interpreting performance. In the present data, both functions which show associations with experience were found to be related to only one or two measures of working memory. More evidence is needed to show how the two working memory functions are related to interpreting and how their development alongside experience affects performance.

If natural cognitive makeup is involved in simultaneous interpreting along the lines suggested above, it would be a plausible link with previous research. In Chapter 3 we discussed several studies comparing working memory and simultaneous interpreting, conducted mostly on non-professional participants (untrained bilinguals, interpreting students). Christoffels (2004) and Tzou et al. (2011) found that both the digit span (an equivalent of the letter span used in the present study) and reading span (similar to the complex span in this study) predicted interpreting accuracy and overall quality of interpreting in untrained participants and beginning and advanced students. Liu (2001) and the present study, on the other hand, did not find any relationship between listening span (Liu's study)/ complex span (present study) and interpreting in professional interpreters. Leaving aside possible methodological flaws, we can offer two plausible explanations. On the one hand, in participant groups with no or limited experience, interpreting performance cannot be explained by highly developed interpreting-specific subskills and must rely on more general cognitive functions. Alternatively, the relationship may not have been detected due to the way simultaneous interpreting was measured. Both Liu and the present study chose specific aspects of simultaneous interpreting: selection of primary over secondary information in the case of Liu's study, and measures of various local processing phenomena in the present study. Christoffels and Tzou, on the other hand, scored the output holistically and without discrimination between several possible processes to achieve the same outcome. Tzou (2011:84) observed "I have discovered (…) that two participants may have equivalent scores on either the [simultaneous interpreting-overall accuracy] or [simultaneous interpreting-accuracy of selected segments] measure but their output would be completely different on the word-to-word and/or sentence-to-sentence level". In other words, we cannot exclude the possibility that complex span tasks would predict performance in experienced interpreters as well. On the basis of the present data, however, we conclude that like previous research, our study has established a link between simultaneous interpreting and working memory, but in contrast to previous research we have found stronger evidence for a relationship of simultaneous interpreting to central executive functions, while research conducted with less-skilled participant groups reported associations with working memory *capacity*, i.e. storage functions.

Previous research also extensively tested the hypothesis of interpreters' superior working memory in comparison with non-interpreters. Our study has not addressed this question directly, but the data provide some indirect evidence. We have already discussed in detail that our data do not support the claim that interpreters' working memory, as measured by complex span tasks, improves with experience. The relationship found in this study between experience and the complex span task was negative (although this was partly driven by age), indicating performance was worse with increasing experience. As for research into the central executive in previous research, only Köpke and Nespoulous (2006) tested a central executive function, using the Stroop test, a measure of ability to inhibit automatic responses. The authors concluded that there was no difference between interpreters and controls on this test. A similar test was included in the present battery of working memory tasks as well – the antisaccade task also measures inhibition of automatic responses – and like Köpke and Nespoulous we did not find any evidence for a relationship between this cognitive function and interpreting.

On a theoretical level, Cowan (2000/1) suggests that the ability to ignore distractions is behind the findings in delayed auditory feedback studies, where interpreters were found to be less affected by listening to a delayed playback of their own voice than students (Moser-Mercer et al., 2000, Fabbro and Darò, 1995). Ability to ignore distractors was found to be related to experience in the present study, and this empirical finding is in line with Cowan's suggestion. This ability may also explain better performance on tasks involving articulatory suppression (see Chapter 3 for a review), including correlation between simultaneous interpreting and recall under articulatory suppression in untrained bilinguals (Christoffels, 2004).

Finally, let us briefly consider the present findings in the wider context of interpreting models. In Chapter 2, we discussed three models of simultaneous interpreting and their treatment of working memory. We pointed out that in two of the models, Gerver's (1976) and Darò and Fabbro's (1994) models, working memory plays the role of a passive store. In the third model by Moser (1978), working memory is involved to a much greater, although underspecified, extent, yet the author explicitly states that it is identical to short-term memory. Our data do not easily fit into any of the models, because the main patterns favour a central-executive involvement, rather than reliance on storage. Let us therefore consider other process models that could better accommodate our findings. Mizuno (2005) put forward a proposal for using Cowan's model of working memory (see Chapter 1) as a basic conceptual framework for simultaneous interpreting. Cowan places emphasis on attentional control and activation of items in working memory, with both time and capacity limitations. Successful interpretation then also means rapid processing of items to avoid overload (Cowan, 2000/1). The concept of attentional control, in the sense of active process management in order to avoid overload, is also central to Gile's Efforts Model (1995), as pointed out by Mizuno (2005). To be fair, one needs to recognise that both Cowan's and Gile's models are more recent, and that both are largely open and underspecified, which makes it easier to accommodate a whole range of findings (see for example Pym, 2009). Nevertheless, both models provide a useful conceptual framework in which the present findings can be accommodated.

6.3. Methodological contributions

Exploratory research by definition charts unknown territory, which limits its capacity for making firm statements and coming to definitive conclusions. Nevertheless, the broadly designed study presented in this dissertation has proved to be a useful way of considering a complex question and finding, if not answers, then at least interesting clues for further, more focused research. In many areas of interpreting research, experimental designs are still premature, with hypotheses derived from intuition rather than a sound theoretical or empirical basis. We believe that designs similar to the one used in this study would be very useful in interpreting research, as they would provide empirical evidence for phenomena that so far were only assumed or hypothesised. In this way, our findings could serve as a basis for hypotheses generation and theory building.

This study also contributes to a better understanding of the use of cognitive psychological methods in interpreting research. Our method relies on basic cognitive research in two ways. First, we used common tests of working memory capacity and executive functions. These tests are stand-alone, i.e. administered outside the context of an interpreting task. They proved to be useful and informative in measuring interpreters' working memory, despite the fact, as is sometimes argued, that they do not resemble interpreting in any way. We believe the opposite is true. By using simple abstract tests, it is easier to define which cognitive function is tested. If relationships with interpreting are found, then the explanation for the association between a complex bilingual task and deciding on the shape and direction of arrows is likely to be in the only component they have in common: the cognitive function which was tested. Relating two complex tasks to each other makes the interpretation of a relationship more difficult, because any number of processes could be involved.

Secondly, we used several cognitive psychological findings to guide our choice of measures of simultaneous interpreting. One such general finding was that individuals with smaller working memory capacity found it more difficult to relate subject and main verb in sentences where these were separated by relative clauses. This syntactic manipulation was embedded in our testing materials, but did not yield any findings. One possible reason is that this finding is usually generated when participants are asked to comprehend such sentences in a monolingual setting and are asked to indicate the subject-verb agreement by a simple yes-no response. In contrast, we expected a similar finding after asking participants to comprehend such sentences in a bilingual setting under conditions of time pressure and dual tasking, and inferred their response from their active verbal production in another language. With hindsight, such specific predictions based on research conducted in a completely different situation are perhaps unrealistically optimistic. Nevertheless, we conclude that the application of cognitive psychological methods in interpreting research can yield interesting results.

6.4. Suggestions for further research

More research will be needed to provide a clearer picture, and some suggestions were made as the individual results and findings were discussed. The first step should be to verify the present findings on an independent data set, although probably with similar characteristics with regard to the participants and materials used in this study. Specifically, our interpreters were fairly experienced, they interpreted between languages which are not a textbook example of linguistic divergence and they performed the task in as ecological a manner as possible. The materials used were intentionally created in a way to challenge even very experienced interpreters, favouring fast and efficient processing over exercises in style and linguistic creativity.

Among the intriguing paths to take is the exploration of individual interpreting styles and their relation to individual cognitive makeup. The logical next step is to compare interpreters on their cognitive abilities and relate them to interpreting styles. In the present study for example, performance on the updating task (2-back) was related to interpreting numbers. Specifically, interpreters who were able to update their memory more efficiently also interpreted more figures accurately, with figures having been scored as accurate only when there was *a perfect match between source and interpreting*. Other styles of interpreting figures are known and often used, such as rounding (*46.5%* interpreted as *46%*), descriptive translation (*about one half*), various approximations (*more than 40%*) and also strategic omissions (Cheung, 2009). Another scoring approach could focus on the number and type of errors. The different types of observed behaviour may in principle reflect different underlying processes and also tap different cognitive functions.

Another very interesting question mark arising from this study is the observed link between natural cognitive functions unrelated to experience and interpreting performance. If working memory is indeed involved in interpreting at all levels, from untrained bilinguals to seasoned professionals, it would be worthwhile to study these relationships in greater detail and explore their potential for candidate selection in interpreting schools. Similarly, and linked to the above suggestion, knowing one’s preferred interpreting style might give interpreter trainers new tools by enabling them to provide more individual advice on interpreting strategies a student may adopt.

References

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General*, *117* (3), 288-318.

Anderson, L. (1994). Simultaneous interpretation: Contextual and translation aspects. In S. Lambert & B. Moser-Mercer (Eds.), *Bridging the gap: Empirical research in simultaneous interpretation* (pp.101-120). John Benjamins: Amsterdam and Philadelphia.

Andrews, G., Birney, D., & Halford, G. S. (2006). Relational processing and working memory capacity in comprehension of relative clause sentences. *Memory & Cognition 34*(6), 1325-40.

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

Baddeley, A. D. (1990). *Human Memory. Theory and Practice*. London: Lawrence Erlbaum Associates.

Baddeley, A. (1996). Exploring the central executive. *The Quarterly Journal of Experimental Psychology 49*A(1), 5-28.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences 4*(11), 417-423.

Baddeley, A. D. (2002). Is working memory still working? *European Psychologist 7*(2), 85-97.

Baddeley, A., Gathercole, S. & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review 105*(1), 158-173.

Baddeley, A. D. & Hitch, G. J. (1974). Working memory. In G. Bower (Ed). *The psychology of learning and motivation: Advances in research and theory*, *Vol. 8*, (pp. 47-89). New York: Academic Press.

Baddeley, A. D. & Logie, R. H. (1999). Working memory: The multiple-component model. In A. Miyake and P. Shah (Eds). *Models of working memory. Mechanisms of active maintenance and executive control* (pp. 28-61). Cambridge: Cambridge University Press.

Baddeley, A. D., Thomson, N. & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior 14,* 575-589.

Bajo, M. T., Padilla, F. & Padilla, P. (2000). Comprehension processes in simultaneous interpreting. In A. Chesterman, N. Gallardo San Salvador & Y. Gambier (Eds). *Translation in Context* (pp. 127-142). Amsterdam and Philadelphia: John Benjamins.

Bayliss, D. M., Jarrold, C., Gunn, D. M. and Baddeley, A. D. (2003). The complexities of complex span: Explaining individual differences in working memory in children and adults. *Journal of Experimental Psychology: General 132*(1), 71-92.

Berch, D. B., Krikorian, R. & Huha, E. M. (1998). The Corsi block-tapping task: methodological and theoretical considerations. *Brain and Cognition*, *38*, 317-338

Bialystok, E. & Craik, F. I. M. (2010). Cognitive and linguistic processing in the bilingual mind. *Current Directions in Psychological Science 19*(1), 19-23.

Buehner, M., Krumm, S. & Pick, M. (2005). Reasoning is working memory is not attention. *Intelligence 33,* 251-272.

Büllow-Møller, A. M. (1999). Existential problems. On the processing of irrealis in simultaneous interpreting. *Interpreting 4*(2), 145-168.

Caplan, D. & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences 22,* 77-126.

Carpenter, P. A., Just, M. A. & Reichle, E. D. (2000). Working memory and executive function: Evidence from neuroimaging. *Current Opinion in Neurobiology 10,* 195–99.

Carpenter, P. A., Miyake, A. & Just, M. A. (1995). Language comprehension: Sentence and discourse processing. *Annual Review of Psychology 46,* 91-120.

Cattell, R. B. & Cattell, A. K. S. (1950). *Test of g: Culture free*. Champaign, IL: Institute for personality and ability testing.

Cheung, A. K. (2009). Numbers in simultaneous interpreting: An experimental study. *Forum*, *7*(2), 61-88.

Chincotta, D. & Underwood, G. (1998). Simultaneous interpreters and the effect of concurrent articulation on immediate memory. A bilingual digit span study. *Interpreting 3*(1), 1-20.

Clark, H. H. (1969). Linguistic processes in deductive reasoning. *Psychological Review 76,* 387-404.

Christoffels, I. (2004). *Cognitive Studies in Simultaneous Interpreting.* Ipskamp/Enschede: PrintPartners.

Christoffels, I. K., de Groot, A. M. B., & Kroll, J. F. (2006). Memory and language skills in simultaneous interpreters: The role of expertise and language proficiency. *Journal of Memory and Language*, *54*(3), 324-345. doi: 10.1016/j.jml.2005.12.004.

Christoffels, I.K, de Groot, A.M.B. & Waldorp, L. J. (2003). Basic skills in a complex task: A graphical model relating lexical retrieval, working memory, and simultaneous interpreting. *Bilingualism: Language and Cognition 6*, 201-211.

Colom, R., Rebollo, I., Abad, F. J. & Shih, P. C.. (2006). Complex span tasks, simple span tasks, and cognitive abilities: A reanalysis of key studies. *Memory and Cognition 34*(1), 158-171.

Colomé, À. (2001). Lexical activation in bilinguals' speech production: Language-specific or language-independent?. *Journal of Memory and Language 45,* 721-736.

Conway, A. R.A., Cowan, N., Bunting, M. F., Therriault, D. J. & Minkoff, S. R.B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence 30,* 163-183.

Conway, A. R.A. & Engle, R. W. (1994). Working memory and retrieval: A resource-dependent inhibition model. *Journal of Experimental Psychology: General 123*(4), 354-373.

Conway, A. R.A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O. & Engle, R. W. (2005). Working memory span tasks: A methodological review and a user's guide. *Psychonomic Bulletin and Review 12*(5), 769-786.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin 104*(2): 163-191.

Cowan, N. (1995). *Attention and Memory. An Integrated Framework.* New York/Oxford: Oxford University Press.

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.) *Models of working memory. Mechanisms of active maintenance and executive control* (pp. 62-101). Cambridge: Cambridge University Press.

Cowan, N. (2000/01). Processing limits of selective attention and working memory: Potential implications for interpreting. *Interpreting, 5*(2), 117-146. doi: 10.1075/intp.5.2.05cow.

Cowan, N. (2005). *Working memory capacity*. New York and Hove: Psychology Press.

Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismajatullina, A. & Conway, A. R. A. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology 51,* 42-100.

Daneman, M. & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior 19,* 450-466.

Daneman, M. & Carpenter, P. A. (1983). Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology: Learning, Memory, and Cognition 9*(4), 561-584.

Daneman, M. & Green, I. (1986). Individual differences in comprehending and producing words in context. *Journal of Memory and Language 25,* 1-18.

Darò, V. (1989). The role of memory and attention in simultaneous interpretation: A neurolinguistic approach. *The Interpreters' Newsletter 2,* 50-56.

Darò, V. & Fabbro, F. (1994). Verbal memory during simultaneous interpretation: Effects of phonological interference. *Applied Linguistics 15*(4), 365-381.

de Groot, A. M. B. (1997). The cognitive study of translation and interpretation. Three approaches. In J. H. Danks, G. M. Shreve, S.B. Fountain & M. K. McBeath (Eds.), *Cognitive processes in translation and interpretation* (pp.25-56). Thousand Oaks, London and New Delhi: Sage.

Déjean le Féal, K. (1982). Why impromptu speech is easy to understand. In N. E. Enkvist (Ed.), *Impromptu speech: A symposium* (221-239). Åbo: Åbo Akademi.

Duff, S. C. & Logie, R. H. (2001). Processing and storage in working memory span. *The Quarterly Journal of Experimental Psychology 54A*(1), 31-48.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science 11*(1), 19-23.

Engle, R. W. & Conway, A. R.A. (1998). Working memory and comprehension. In R.H. Logie & K.J. Gilhooly (Eds.), *Working Memory and Thinking* (pp. 67-91). Hove: Psychology Press.

Engle, R. W. & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory, Vol.44* (pp. 145-199). New York: Academic Press.

Engle, R. W., Tuholski, S. W., Laughlin, J. E. & Conway, A. R.A (1999). Working memory, short-term memory, and general fulid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General 128*(3), 309-331.

Ericsson, K. A. (2000). Expertise in interpreting. An expert-performance perspective. *Interpreting 5*(2), 187-220.

Ericsson, K. A. & Delaney, P. F. (1998). Working memory and expert performance. In R.H. Logie & K.J. Gilhooly (Eds.), *Working memory and thinking* (pp. 93-114). Hove: Psychology Press.

Ericsson, K. A. & Delaney, P. F. (1999). Long-term working memory as an alternative to capacity models of working memory in everyday skilled performance. In A. Miyake & P. Shah (Eds.), *Models of working memory. Mechanisms of active maintenance and executive control* (pp. 257-297). Cambridge: Cambridge University Press.

Ericsson, K. A. & Kintsch, W. (1995). Long-term working memory. *Psychological Review 102*(2), 211-245.

Ericsson, K. A. & Simon, H. (1980). Verbal reports as data. *Psychological Review*, *87*, 215-251.

Ericsson, K. A. & Simon, H. A. (1993). *Protocol analysis: verbal reports as data*. Cambridge, MA: MIT Press.

Eriksen, B. A. & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception and Psychophysics, 16*(1), 143-149.

Fabbro, F., & Darò, V. (1995). Delayed auditory feedback in polyglot simultaneous interpreters. *Brain and language*, *48*, 309-319.

Fan, J., McCandliss, B.D., Sommer, T., Raz, A. & Posner, M.I. (2002) testing the efficiency and independence of attentional networks, *Journal of Cognitive Neuroscience, 14*(3), 340-347.

Feldman Barrett, L., Tugade, M. M. & Engle, R. W. (2004). Individual differences in working memory capacity and dual-process theories of the mind. *Psychological Bulletin 130*(4), 553-573.

Frauenfelder, U. H. & Schriefers, H. (1997). A psycholinguistic perspective on simultaneous interpretation. *Interpreting 2*(1/2), 55-89.

Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: a latent-variable analysis. *Journal of experimental psychology. General*, *133*(1), 101-35. doi: 10.1037/0096-3445.133.1.101.

Gathercole, S. E. & Baddeley, A. D. (1993). Phonological working memory – a critical building-block for reading development and vocabulary acquisition. *European Journal of Psychology of Education 8*(3), 259-272.

Gernsbacher, M. A. & Foertsch, J. A. (1999). Three models of discourse comprehension. In S. Garrod & M.J. Pickering (Eds.), *Language processing* (pp. 283-299). Hove: Psychology Press.

Gerver, D. (1969/2002). The effects of source language presentation rate on the performance of simultaneous conference interpreters. In F. Pöchhacker & M. Shlesinger (Eds.), *The Interpreting Studies Reader* (pp. 52-66). Routledge: London and New York.

Gerver, D. (1974). The effects of noise on the performance of simultaneous interpreters: Accuracy of performance. *Acta Psychologica 38*, 159‐167.

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

Gerver, D. (1976). Empirical studies of simultaneous interpretation: A review and a model. In R.W. Brislin (Ed.), *Translation. application and research* (pp. 165-207). New York: Garden Press.

Gile, D. (1995). *Basic concepts and models for interperter and translator training.* Amsterdam and Philadelphia: John Benjamins.

Gile, D. (1997). Conference interpreting as a cognitive management problem. In J. H. Danks, G. M. Shreve, S. B. Fountain, & M. K. McBeath (Eds.), *Cognitive processes in translation and interpreting* (pp. 196-214). Thousand Oaks, London and New Delhi: Sage.

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

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition 1*, 67-81.

Hambrick, D. Z. & Oswald, F. L. (2005). Does domain knowledge moderate involvement of working memory capacity in higher-level cognition? A test of three models. *Journal of Memory and Language 52*, 377-397.

Hermans, D., van Dijk, R., & Christoffels, I. (2007). De effectiviteit van gebarentaaltolken in de communicatie tussen dove en horende mensen [Effectiveness of sign language interpreting in communication between deaf and hearing people]. Unpublished final project report. Utrecht University of Applied Sciences, the Netherlands.

Hester, R. & Garavan, H. (2005). Working memory and executive function: The influence of content and load on the control of attention. *Memory and Cognition 33*(2): 221-233:

Hitch, G. J. (2005). Working memory. In N. Braisby & A. Gellatly (Eds.), *Cognitive Psychology* (pp. 307-341). Oxford: Oxford University Press.

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

Howell, D. C. (2010). *Statistical methods for psychology. International edition* (7th ed.). Belmont: Wadsworth.

Hunt, E. (1978). Mechanics of verbal ability. *Psychological Review 85*(2), 109-130.

Iman, R. L. & Conover, W.J. (1979). The use of the rank transform in regression. *Technometrics, 21*(4), 499-509.

Ivanova, A. (1999). *Discourse processing during simultaneous interpreting. An expertise approach*. Unpublished doctoral dissertation, University of Cambridge.

James, W. (1890). *The Principles of Psychology*. New York: Henry Holt.

Jarrold, C. & Towse, J. N. (2006). Individual differences in working memory. *Neuroscience 139*, 39-50.

Jarvella, R. J. (1971). Syntactic processing of connected speech. *Journal of Verbal Learning and Verbal Behavior 10*, 409-416.

Just, M. A. & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review 99*(1), 122-149.

Kapranov, O. (2008). *The Impact of Language Exposure on Fluency in Simultaneous Interpreting*. Unpublished PhD dissertation, University of Western Australia. Retrieved from https://repository.uwa.edu.au/ in June 2011.

Kiewra, K. A. & Benton, S. L. (1988). The relationship between information-processing ability and notetaking. *Contemporary Educational Psychology 13*, 33-44.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language 30*, 580–602.

Kintsch, W. (1998). *Comprehension. A paradigm for cognition*. Cambridge: Cambridge University Press.

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

Köpke, B. & Signorelli, T. M. (2011). Methodological aspects of working memory assessment in simultaneous interpreters. *International Journal of Bilingualism.* Advance online publication. doi: 10.1177/1367006911402981.

Kroll, J. F., Bobb, S. C., Misra, M. & Guo, T. (2008). Language selection in bilingual speech: Evidence for inhibitory processes. *Acta Psychologica 128,* 416-430.

Kyllonen, P. C. & Christal, R. E. (1990). Reasoning ability is little more than working memory capacity? *Intelligence 14,* 389-433.

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

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

Lee, T.-H. (2002). Ear-voice span in English into Korean simultaneous interpretation. *Meta*, *47*(4), 596-606.

Lépine, R., Bernardin, S. & Barrouillet, P. (2005). Attention switching and working memory spans. *European Journal of Cognitive Psychology 17*(3), 329-345.

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

Liu, M., Schallert, D. L., & Carroll, P. J. (2004). Working memory and expertise in simultaneous interpreting. *Interpreting, 6*(1), 19-42. doi: 10.1075/intp.6.1.04liu.

Logie, R. H. (2006). Testing the world's working memory. Invited talk at the University of Ghent, Ghent, Belgium, 20th November 2006.

Massaro, D. W. & Shlesinger, M. (1997). Information processing and a computational approach to the study of simultaneous interpretation. *Interpreting 2*(1/2), 13-53.

Matthews, G., Davies, D. R., Westerman, S. J. & Stammers, R. B. (2000). *Human performance. cognition, stress and individual differences.* Hove: Psychology Press.

Meinz, E. J. & Hambrick, D. Z. (2010). Practice is necessary but not sufficient to explain individual differences in piano sight-reading skill: The role of working memory capacity. *Psychological Science 21*(7), 914-919.

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.

Miyake, A. & Friedman, N. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In A. Healy & L. Bourne (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 339-365). London: Lawrence Erlbaum Associates.

Miyake, A., Just, M. A. & Carpenter, P. A. (1994). Working memory constraints on the resolution of lexical ambiguity: Maintaining multiple interpretations in neural contexts. *Journal of Memory and Language 33*(2), 175-202.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A. & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41,* 49-100.

Mizuno, A. (2005). Process model for simultaneous interpreting and working memory. *Meta 50*(2): 739-752.

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

Moser-Mercer, B. (1997a). Beyond curiosity: Can interpreting research meet the challenge? In J. H. Danks, G. M. Shreve, S. B. Fountain, & M. K. McBeath (Eds.), *Cognitive processes in translation and interpreting* (pp. 176-205). Thousand Oaks, London and New Delhi: Sage.

Moser-Mercer, B. (1997b). The expert-novice paradigm in interpreting research. In E. Fleischmann. In W. Kutz & P. A. Schmitt (Eds.), *Translationsdidaktik. Grundfragen der Übersetzungswissenschaft* (pp. 255-261). Tübingen: Gunter Narr.

Moser-Mercer, B. (2005). Simultaneous interpreting and cognitive limitations. The acquisition of expertise as a process of circumventing constraints. Keynote delivered at the Workshop on “Cognitive aspects of simultaneous interpreting”, 19-20 May 2005, Université de Toulouse-Le Mirail.

Moser-Mercer, B., Frauenfelder, U. H., Casado, B., & Künzli, A. (2000). Searching to define expertise in interpreting. In B. Englund Dimitrova & K. Hyltenstam (Eds.), *Language processing and simultaneous interpreting* (pp. 107-131). Amsterdam and Philadelphia: John Benjamins.

Naveh-Benjamin, M. & Ayres, T. J. (1986). Digit span, reading rate, and linguistic relativity. *Quarterly Journal of Experimental Psychology*, *38A*, 739-751.

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

Oberauer, K. (2002). Access to information in working memory: Exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition 28*(3), 411-421.

Oberauer, K. (2006). Is the focus of attention in working memory expanded through practice? *Journal of Experimental Psychology: Learning, Memory and Cognition 32*(2), 197-214.

Oberauer, K. & Göthe, K. (2006). Dual-task effects in working memory: Interference between two processing tasks, between two memory demands, and between storage and processing. *European Journal of Cognitive Psychology 18*(4), 493-519.

Oberauer, K., Süss, H.-M., Wilhelm, O. & Wittman, W. W. (2003). The multiple faces of working memory: Storage, processing, supervision and coordination. *Intelligence 31*, 167-193.

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

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

Petite, C. (2003). Repairs in simultaneous interpreting: quality improvement or simple error correction? In A. Collados Aís, M. M. Fernández Sánchez & D. Gile (Eds.), *La evaluación de la calidad en interpretación: investigación* [Quality assessment in interpreting: research] (pp. 61-71). Granada: Comares.

Pöchhacker, F. (2004). *Introducing Interpreting Studies*. London and New York: Routledge.

Pym, A. (2009). On omission in simultaneous interpreting: Risk analysis of a hidden effort. In G. Hansen, A. Chesterman & H. Gerzymisch-Arbogast (Eds.), *Efforts and Models in Interpreting and Translation Research* (83-105). Amsterdam and Philadelphia: John Benjamins.

Ranganath, C. & Blumenfeld, R. S. (2005). Doubts about double dissociations between short- and long-term memory. *Trends in Cognitive Sciences 9*(8), 374-380.

Redick, T. S. & Engle, R. W. (2006). Working memory capacity and attention network test performance. *Applied Cognitive Psychology 20,* 713-721.

Repovš, G. & Baddeley, A. (2006). The multi-component model of working memory: Explorations in experimental cognitive psychology. *Neuroscience 139,* 5-21.

Rogers, R. D. & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General, 124*(2), 207-231.

Ruchkin, D. S., Grafman, J., Cameron, K. & Berndt, R. S. (2003). Working memory retention systems: A state of activated long-term memory. *Behavioral and Brain Sciences 26,* 709-777.

Seleskovitch, D. (1968/1978 English). *Interpreting for international conferences. problems of language and communication.* Washington: Pen and Booth.

Seleskovitch, D. (1975/2002). Language and memory: A study of note-taking in consecutive interpreting. In F. Pöchhacker & M. Shlesinger (Eds.), *The Interpreting Studies reader* (pp. 121-129). London and New York: Routledge.

Setton, R. (1999). *A cognitive-pragmatic analysis of simultaneous interpretation*. Amsterdam and Philadelphia: John Benjamins.

Setton, R. (2001). Deconstructing SI: A contribution to the debate on component processes. *Interpreters' Newsletter 11*, 1-26.

Shlesinger, M. (2000). *Strategic Allocation of Working Memory and Other Attentional Resources in Simultaneous Interpreting*. Unpublished PhD dissertation, Bar Ilan University, Ramat Gan, Israel.

Signorelli, T. M., Haarmann, H. J. & Obler, L. K. (2011). Working memory in simultaneous interpreters: effects of task and age. *International Journal of Bilingualism*. Advance online publication. doi: 10.1177/1367006911403200.

Sohn, Y. W. & Doane, S. M. (2003). Roles of working memory capacity and long-term working memory skill in complex task performance. *Memory and Cognition 31*(3), 458-466.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*, 643-662.

Swets, B., Desmet, T., Hambrick D. Z. & Ferreira, F. (2007). The role of working memory in syntactic ambiguity resolution: A psychometric approach. *Journal of Experimental Psychology: General 136*(1), 64-81.

Timarová, Š. (2007). *Measuring working memory in interpreters*. Unpublished DEA thesis, ETI, University of Geneva.

Timarová, Š. (2008). Working memory and simultaneous interpreting. In P. Boulogne (Ed.), *Translation and its others. Selected papers of the CETRA research seminar in Translation Studies 2007.* Available from http://www2.arts.kuleuven.be/info/bestanden-div/Timarova.pdf.

Timarová, Š., Dragsted, B. & Hansen, I. G. (2011). Time lag in translation and interpreting. A methodological exploration. In C. Alvstad, A. Hild & E. Tiselius (Eds.), *Methods and strategies of process research. Integrative approaches in Translation Studies* (pp. 121-146). Amsterdam and Philadelphia: John Benjamins.

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.

Turner, M. L. & Engle, R. W. (1989). Is working memory capacity task dependent?. *Journal of Memory and Language 28,* 127-154.

Tzou, Y.-Z. (2008). *The roles of working memory, language proficiency, and training in simultaneous interpretation performance: Evidence from Chinese-English bilinguals*. Unpublished PhD dissertation. Texas A&M University, Texas.

Tzou, Y.-Z., Eslami, Z. R., Chen, H.-C. & Vaid, J. (2011). Effects of language proficiency and degree of formal training in simultaneous interpreting on working memory and interpreting performance: evidence from Mandarin-English speakers. *International Journal of Bilingualism*. Advance online publication. doi: 10.1177/1367006911403197.

Vandierendonck, A., Kemps, E., Fastame, M.C. & Szmalec, A. (2004). Working memory components of the Corsi block task. *British Journal of Psychology, 95,* 57-79.

Was, C. A. & Woltz, D. J. (2007). Reexamining the relationship between working memory and comprehension: The role of available long-term memory. *Journal of Memory and Language 56,* 86-102.

Waters, G S. & Caplan, D. (1996). The measurement of verbal working memory capacity and its relation to reading comprehension. *The Quarterly Journal of Experimental Psychology 49A*(1), 51-79.

Waters, G. & Caplan, D. (2005). The relationship between age, processing speed, working memory capacity, and language comprehension. *Memory 13*(3/4),403-413.

Woltz, D J. (1988). An investigation of the role of working memory in procedural skill acquisition. *Journal of Experimental Psychology: General 117*(3), 319-331.

Software

Adobe Systems Incorporated (2003). Adobe Audition (Version 1.0) [Computer software] Available from www.adobe.com

Anthony, L. (2011). AntConc (Version 3.2.4) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.antlab.sci.waseda.ac.jp/

Dialang. Language diagnostic software. Available at http://www.lancs.ac.uk/researchenterprise/dialang/about (previously www.dialang.org)

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39,* 175-191.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime User's Guide*. Pittsburgh: Psychology Software Tools Inc.

Annexes

Annex 1: Statistical concepts

M, mean	the average of a set of scores
SD, standard deviation	the average difference between a mean and scores on which it was calculated. For example, if two interpreters have two working languages each, the mean will be two languages. If two other interpreters have one and three working languages, the mean will also be two languages. The standard deviation is a measure of how the scores from which the mean was calculated differ from the mean. In the first case, the standard deviation is zero. In the second, it is equal to one.
median	the middle value in a set of scores. If the set contains 5 scores, the third value in the ordered set is the median.
skewness	a measure of symmetry in the distribution of scores. Skewed distributions are asymmetrical.
kurtosis	a measure of flatness or thinness of the distribution of scores.
reliability	a measure of how consistent the test is. In a reliable test, responses to all equivalent items will yield similar results.
correlation	a measure of relationship in two variables. Two variables are positively related if increase in one is associated with increase in the other (for example age and height in children) or if increase in one is associated with decrease in the other (for example age and physical fitness). Positive and negative relationships are a matter of interpretation. For example, as the temperature decreases, people wear increasingly more clothes (negative relationship), or as the cold increases, people wear more clothes (positive relationship), although both describe the same reality.
Pearson correlation coefficient	a way to quantify correlation. The values range from -1 (perfect negative correlation) through 0 (no relationship) to +1 (perfect positive correlation). The further away the value of the coefficient is from zero (towards +1 or -1), the stronger the relationship is.
Spearman correlation coefficient	another way to quantify correlation. Instead of using the actual values, the scores are first ranked. Pearson correlation coefficient is then calculated on the ranks. Spearman is a method useful in small samples.
ranked data	data ordered from low to high values and assigned a rank
p, significance probability	the *p* value is a numerical estimation of how likely it would be to observe a given relationship if in fact the relationship did not exist and it was only observed due to a coincidence in the data. By convention, values of *p* smaller than 0.05 are considered to reflect a very small chance that the observed relationship is a coincidence. When *p* value is below 0.05, we conclude that the observed relationship is real.

regression	similar to correlation, regression uses one variable to predict the outcome of another variable.
multiple linear regression	the outcome of one variable is predicted on the basis of several variables (predictors)

Annex 2: Simultaneous interpreting. Text Amnesty

Text as delivered. Manipulated sentences are marked in bold

Thank you, Robin. It is a pleasure, as always, to have the opportunity to speak at one of your seminars. Let me introduce the topic of my speech by several real stories.

On April 27 2005, Shi Tao, a Chinese journalist, received a ten-year prison term for sending information about a Communist Party decision to a website in the US, using his Yahoo email account. On June 2, Shi's appeal was denied. He was imprisoned solely for peacefully exercising his legitimate right to seek, receive and impart information. In the eyes of Amnesty International, Shi is a prisoner of conscience.

The evidence which led to Shi's conviction and which was provided by Yahoo! Hong Kong, included details of his Yahoo mail account and the IP address, despite a confidentiality agreement between Shi and Yahoo!. The company also confirmed the precise time when Shi wrote his email.

This is but one example of how corporations participate in human rights cases. Let me give you another example, an incident which happened in Nigeria in 2005. **Nigerian soldiers, who served in special military forces and who were heavily armed with automatic rifles, used live ammunition against protesters from a small ethnic community for entering Chevron's oil terminal.** The encounter was bloody. Chevron Nigeria, which operates the terminal, said that several employees and security officers received minor injuries. One demonstrator was shot and later died from his injuries. **At least 30 others were injured, 15 of them seriously, by blows from rifle butts and other weapons.** There was also material damage. **The boundary fence was cut in 35 places, and 87 windows and 17 helicopter windscreens were smashed.** It was several hours before the injured protesters could reach a hospital, a lengthy boat journey away. Neither the security forces nor Chevron Nigeria provided adequate medical care or assistance to transport the injured.

The protest was over a Memorandum of Understanding signed by community representatives and Chevron Nigeria in 2002. **The protesters said that Chevron Nigeria had not provided the 1750 jobs and 55 development projects they were promised.** The company denied all charges. It said that the responsibility for protecting its facilities rests with the state security forces, and it could not control the actions of the security forces in any way. Chevron, like other oil companies operating in Nigeria, does provide the security forces with oil in line with industry practice. **The company did not promise not to assist the security forces in the future.**

Human rights law is clear. The primary responsibility for respecting and promoting human rights rests with the state. Human rights law is also clear on other issues. **Non-state actors which include privately-owned companies and which are considered an organ of society, must uphold human rights within their area of influence in the context of both their operations and local communities.** But what happens when the company itself does not commit an abuse but benefits from an abuse committed by a government or armed group? **Or the company does not disaffiliate itself from such governments or armed groups?** Or it funds those who commit abuses? Or complies with national laws and policies which are clearly in violation of international human rights?

The UN calls upon companies not to be complicit in human rights abuses. UN Human Rights guideline specifically states (and I quote): "Transnational corporations and other business enterprises shall have the responsibility to use due diligence in ensuring that their activities do not **contribute** directly or indirectly to human abuses, and that they do not directly or indirectly **benefit** from abuses of which they were aware or ought to have been **aware**….. Transnational corporations and other business enterprises shall inform themselves of the human rights, impact of their principal activities and major proposed activities so that they can further avoid **complicity** in human rights abuses." End of quote.

This kind of involvement of companies is known as corporate complicity. **Corporate complicity, which is now attracting attention of the international community and which is an emerging area of law, poses important moral and legal questions about borderlines between right and wrong.** There are many different situations. On the one hand, companies may have knowingly funded, supported or benefited from human rights abuse**. On the other, companies simply did not disapprove of abuse committed by others.** Where do the boundaries of complicity begin and end? On the one side there is the law which tells us what can and cannot be done. But we must not stop where the law rests. **Values and principles which often reach beyond existing law and which are worth fighting for should be the basis of our efforts for improved legislation in an international context.** To give an analogy, there is currently no international consensus on the abolition of the death penalty. Amnesty International continues to work towards its abolition, and has been instrumental in setting the international strategy for it. **But many countries do not undertake not to apply the penalty.**

In some cases, criminal or civil liability may be clear. For instance there is jurisprudence on complicity in war crimes, crimes against humanity and genocide. But there are many other cases where the definitions of complicity are only emerging. This sometimes happens through campaigning pressure. At other times, the rules emerge through greater acknowledgement of corporate responsibility. Occasionally, we learn hard lessons from disasters and failures. **Yet, many companies do not disapprove of a number of practices.** The reason is that they often don't know where to draw the line. So, let me start by laying out some examples of real cases where companies can risk complicity.

In the area of <u>Armed Conflict:</u>

Companies may provide money, resources, infrastructure, products or services that facilitated human rights violations in the context of armed conflict. The support is not necessarily direct. **Some analysts estimate that 13 to 15 of every 100 companies present in zones of conflict may indirectly contribute to armed conflicts.** An apparently innocuous trade in rough diamonds was used to fund weapons with which gross human rights abuses were then committed. No court case had to be fought to make the diamond industry realize that the risk to their reputation of complicity in crimes against humanity. That led to the Kimberly Process certification scheme.

I want to move on to the area of <u>Slavery and forced labour:</u>

The right not to be held in slavery or servitude was recognized in Article 2 of the 1926 Slavery Convention. The convention obliges states to prevent slave trade and abolish slavery in all forms. The International Labour Organisation adopted a convention in 1930 to end forced labour. **Companies which use migrant labour directly or which rely on suppliers of workers, must check the identity of their workers very carefully because of an increase in human trafficking.** Migrant workers are especially vulnerable to abuse. When a gang master hires undocumented foreigners to work under inhumane conditions in certain industries, the companies that benefits from the products they make could run the risk of being complicit, even though they themselves do not own the factories.

<u>*(Child Labour)*[31]</u>

Many companies in the developing world do not respect the rule not to employ children. Many of those companies supply their products to major international chains in the developed

[31] Segments in italics and in brackets should have been said but were not. They are kept here as they clarify the text or its structure.

world. Ending child labour will take time, but some forms of child labour are inherently exploitative, dangerous and unacceptable. **A complicit company does not explicitly disapprove of such child labour.** If it continues to contract work to offending subcontractors, it could run the risk of being complicit in the abuse of children's rights.

I'd like to move on now to the area of Technology:

There have also been instances where companies have provided technology to governments to commit human rights abuses, for instance, surveillance technology to authoritarian governments which then used that technology to track down and punish dissidents. The role of international IT companies in China has drawn great attention in recent years. The Chinese government cannot do this on its own; it is able to do so because companies are assisting the government in blocking certain sites and sources of information. Earlier I mentioned the case of Shi Tao – could such a case create a claim of corporate complicity on the part of Yahoo? What about Google and Microsoft? **They have not undertaken not to assist the Chinese governments in restricting freedom of expression.** Such corporate behaviour raises issues of complicity. It also undermines the value of an industry committed to free access of information.

(Discrimination)

Now, at the heart of human rights law is the principle of non-discrimination. **Statistics show that women are 27% less likely, and elderly persons 63% less likely to be hired.** Companies need to be careful about non-discrimination in employment practice, including retirement and promotion. They need to be careful in their choice of suppliers and partners; in meeting quotas under affirmative action programs where such programs are in place. Companies need to be careful also in a whole range of other policies, including laws that specifically discriminate against women and minorities.

Many companies seek to hide behind national laws. But what may be legal at the national level could be wholly unacceptable at the international level. **Many companies, which had operations in South Africa during the apartheid and which followed the laws of that time, effectively aided the South African government in implementing its discriminatory policies.**

Standards for human rights are set internationally, but unfortunately the international legal system is too weak to enforce international standards. But this does not exonerate companies from moral scrutiny and reputational damage. Some of the toughest campaigns against corporate behaviour were not fought in the court of law but in the court of public opinion, as Calito, a textiles manufacturer, found to its detriment. After it became known that the company allowed child labour in its East-Asian factories, a public campaign lead to its boycott by customers. **The company saw its turnover drop by 54% and suffered a USD 127 million loss in just 2 years.**

It is clear from these examples that the concept of complicity is nuanced and multilayered, with different meanings in different contexts. How to tease them apart? **Sometimes, a company simply does not disapprove of a certain practice.** Is it complicit? At one end of the spectrum complicity must be differentiated from direct abuse. There are many cases in which companies are directly responsible for human rights abuses e.g. the disaster at the Union Carbide plant in Bhopal, India. At the other end of the spectrum, the mere presence of a company in a country with a poor human rights record does not make it complicit in human rights abuse.

Criteria for defining corporate complicity are emerging from existing jurisprudence and by analogy with complicity in international and domestic criminal law. And through academic research and analysis, but also through greater consciousness about the role and impact of corporate behaviour on human rights. **The growing body of research now includes 23 academic studies and at least 138 court findings.** It is not easy to understand corporate complicity. In order to do so, it is necessary to know the nature and scale of a company's participation in the economy, the way and extent to which the company supports or benefits from the human rights abuses, the nature and duration of the company's relationship to the perpetrator and the victims, the company's knowledge and intent, and the scope and character of the abuse itself. **Emerging jurisprudence in criminal law, which builds on complicity for crimes against humanity and which has provided a broad definition of complicity, helps develop civil standards and principles in different tribunals and courts.** Key considerations are:

a) <u>Proximity</u>: How close is the company to the violation? The higher proximity, the higher the likelihood of complicity.

b) <u>Knowledge and Awareness</u>: Did the company know, or should have known about the abuse? A large multinational may not always be aware of every detail in every subsidiary. **Nike, the sportswear manufacturer, has 657 suppliers spread across 65 countries, not including the US, where it is headquartered.** Nike initially did not know the conditions in which employees of its sub-contractors worked until activist groups highlighted the abuses. Should it have known? **Importantly, Nike had not explicitly promised not to allow this happen again.** In the end, Nike could not get away from the reputational damage, and had to establish a mechanism to monitor workplace conditions.

<u>Benefit:</u> Does the company benefit from the abuse? Any benefit the company derives as a consequence of the abuses raises the risk of complicity. **A company which wishes to build a golf course and which accepts that the state evicts people from that land, derives a benefit from an abuse of human rights.** A state that passes legislation to ban unions so that companies can operate in free trade zones at lower costs also creates a benefit for the company. In such cases, the company is not actively assisting the state, nor in a partnership with the state, but it stands to benefit from the state's act. This is by far the most common form of corporate complicity. Its particular danger is in the fact that companies may not even be aware they are benefiting from the state's act. **Some estimates suggest that 46% of companies operating in Asia, 51% in South America and 63% in Africa may be in this position.**

<u>(Avoiding complicity)</u>

What can companies do to avoid complicity? What companies must do is to create and maintain a culture of non-complicity in every aspect of their operations. To do so, companies must move towards a culture of compliance with human rights and international standards. For guidance, they can turn to principles suggested by the Office of the High Commissioner on Human Rights. Companies should also seek broader, industry wide commitment to non-complicity. And there are examples of how it could be done. **The Kimberley process, in which companies have joined with governments and NGOs and which established voluntary codes of conduct for the diamond industry, helps avoid complicity in human rights abuses from diamond trade in conflict ridden countries.** The Kimberley Process is not a perfect system, but it shows what can be achieved.

Of course, corporate complicity does not weaken state responsibility – indeed, it strengthens it by ensuring that a perpetrator state is deprived of allies. Just as good citizens avoid being complicit in crime, so good companies must avoid complicity in their sphere of influence.

Thank you.

Annex 3: Simultaneous interpreting. Companies texts - China, Brazil

Text China

Two studies of corporate human rights policies were conducted under the mandate of the United Nations in 2006. These studies provided new information and have been widely referenced. But they also suffered from limitations in geographical coverage. And because of linguistic limitations in the research team at the time, the business recognition report included only Brazilian, Chinese and Russian companies whose sources of information were available in English. The present reports primarily draw on Mandarin and Portuguese sources, along with some English sources.

Our aim in this part of the survey was to map the human rights policies of internationally engaged Chinese companies. Our sample consists of companies in the following sectors: 37 companies from the extractive industry, 12 companies from the financial services industry, 13 companies from the computer hardware industry, 11 companies engaged in transport and logistics, 40 companies engaged in infrastructure and utilities, 20 companies working in the IT and telecommunications sector, 24 companies from the pharmaceutical industry, 10 companies from the chemical industry, and 36 non-classified enterprises.

In the policies of the 203 Chinese companies, we documented a combined total of 672 references to various areas of rights.

Among these Chinese companies, the right to a safe work environment is recognized most frequently (126 firms, or 62.1% of the sample), followed by non-discrimination in employment (111 companies, or 54.7% of the sample), right to physical and mental health (94 companies, or 46.3% of the sample), the right to collective bargaining (89 companies, or 43.8% of the sample), the right to education (78 companies, or 38.4% of the sample), the right to equal pay for equal work (72 companies, or 35.5% of the sample), the right to political life (68 companies, or 33.5% of the sample), and the abolition of forced labor (34 companies, 16.7% of the sample). No other labor-related rights receive recognition from any company.

As noted, the *right to a safe work environment* is recognized by 126 companies in the sample (62.1% of the companies surveyed). This compares to a 75% recognition rate in the survey conducted among US companies. Some companies also refer to specific steps, programs, and reporting mechanisms undertaken to strengthen workplace safety.

The *right to physical and mental health* received similarly frequent recognition—94 of the 203 companies, or 46.3%, compared to 90% of companies in the US report.

The *right to non-discrimination in employment* as well as *the right to equal pay for equal work* are acknowledged by 111 and 72 companies respectively. Most firms mention existing workers' unions and refer to their importance in enabling employees to enforce these principles. 89 companies also address the right to collective bargaining and 68 companies support the right to political life. This is in stark contrast to the US companies surveyed, where more than 90% of companies recognize these rights. A more general *right to education* was recognized by 78 Chinese companies (or 38.4% of the sample).

Finally, 34 companies express support for the abolition of *forced labor* (that's 16.7% of the companies surveyed). In the US report, 60% of companies recognize this prohibition.

No other rights receive recognition from any companies.

As indicated at the outset, further research examining a larger number of Chinese companies is necessary before any definitive conclusions can be drawn about their human rights policies. It is our hope that these ongoing efforts to map company policies will contribute to a deeper understanding among companies and governments alike of patterns and trends in corporate recognition of human rights, and thereby help identify effective means of achieving improved corporate human rights practices throughout the world. Thank you.

Text Brazil

Two studies of corporate human rights policies were conducted under the mandate of the United Nations in 2006. These studies provided new information and have been widely referenced. But they also suffered from limitations in geographical coverage. And because of linguistic limitations in the research team at the time, the business recognition report included only Brazilian, Chinese and Russian companies whose sources of information were available in English. The present reports primarily draw on Mandarin and Portuguese sources, along with some English sources.

Our aim in this part of the survey was to map the human rights policies of internationally engaged Brazilian companies. Our sample consists of companies in the following sectors: 15 companies from the automotive industry, 32 companies from the insurance services industry, 26 companies from the food and beverage industry, 28 companies engaged in forestry and fishing, 14 companies engaged in wholesale and retail, 28 companies working in the media and advertising sector, 30 companies from the construction industry, 18 companies from the legal services industry, and 15 non-classified enterprises.

In the policies of the 206 Brazilian companies, we documented a combined total of 618 references to various areas of rights.

Among these Brazilian companies, the right to work and vocational training is recognized most frequently (103 firms, or 50% of the sample), followed by freedom of association (101 companies, or 49% of the sample), prohibition of degrading treatment (97 companies, or 47.1% of the sample), the right to equal protection under the law (87 companies, or 42.2% of the sample), the right to development (83 companies, or 40.3% of the sample), the right to organize (67 companies, or 32.5% of the sample), equality at work (44 companies, or 21.4% of the sample), and the abolition of child labor (36 companies, 17.5% of the sample). No other labor-related rights receive recognition from any company.

As noted, the *right to work and vocational training* is recognized by 103 companies in the sample (that’s 50% of the sample). This compares to a 75% recognition rate in the survey conducted among US companies. Some companies also refer to specific steps, programs, and reporting mechanisms undertaken to strengthen vocational training.

The *prohibition of degrading treatment* received similarly frequent recognition - 97 of the 206 companies, or 47.1% of the companies surveyed, compared to 90% of companies in the US report.

The *freedom of association* as well as the *right to organize* are acknowledged by 101 and 67 companies respectively. Most firms mention existing workers’ unions and refer to their importance in enabling employees to enforce these principles. 87 companies also address the right to equal protection under the law and 44 companies support equality at work. This is in stark contrast to the US companies surveyed, where more than 90% companies recognize these rights. A more general *right to development* was recognized by 83 Brazilian companies (or 40.3% of the sample).

Finally, 36 companies express support for the abolition of *child labour* (that’s 17.5% of the companies surveyed). In the US report, 60% of companies recognize this prohibition.

No other rights receive recognition from any companies.

As said at the beginning, further research examining a larger number of Brazilian companies is needed before we can draw any definitive conclusions about their human rights policies. We hope that the continuous efforts to survey company policies will contribute to a deeper understanding among companies and governments alike of trends in corporate recognition of human rights, and therefore help identify effective means of improving corporate human rights practices throughout the world. Thank you.

Annex 4: Simultaneous interpreting. Background materials for interpreters

Introduction to text Amnesty

(The following text was video-recorded and played before the interpreting task to provide context)

Good morning, ladies and gentlemen,
welcome to Brussels. And welcome to this seminar on Business and Human Rights. This year this event has taken on a special significance, held as it is just before Human Rights Day. As a matter of fact, it is the third event in a series which we started in 2005, when the Business and Human Rights first conference began to explore the issue of business responsibility and business complicity with regard to human rights abuses. The notion of business responsibility in this area forms the basis of the Second Principle of the United Nations Global Compact, and it also underpins other business and human rights standards, such as for example the OECD Guidelines and the Voluntary Principles on Security and Human Rights. Although the first two conferences provided a lot of fruitful information and a lot of food for thought, there are serious issues that remain to be discussed. For example, how is complicity to be understood? How far do a business's responsibilities extend? What can businesses do to minimise the risk of being complicit in abuses of human rights? These are serious issues indeed. And to debate these issues here, at this conference in Brussels, we are pleased to welcome key figures from a wide variety of international organisations, including many from the business community, as well as from many NGOs. I think we will have an extremely interesting conference, I thank you for being here and I thank you for your interest in these issues. Thank you very much.

Introduction to companies texts

(The following text was video-recorded and played before the interpreting task to provide context)

Ladies and gentlemen,
it is now my pleasure to introduce the next section of our event. You may remember that in 2005, two studies were conducted under the mandate of the United Nations Secretary-General's Special Representative for Business and Human Rights. The studies aimed to map business recognition of human rights by surveying their documents. The two studies reported on a questionnaire survey of the Fortune 100 companies, as well as on a more detailed analysis of the actual human-rights related policies of a broader cross-section of more than 300 firms, as described in their annual reports, their websites and similar public sources. Both these studies were subsequently published and are generally known as the Business Recognition Report. The United Nations Human Rights council considered the effort worthwhile and commissioned more studies, which would focus on companies above and beyond the Fortune list. As a result, several more studies have been carried out and Steven will present two of them during the course of this event. The two presentations will cover Chinese and Brazilian companies.

THE 2009 BUSINESS AND HUMAN RIGHTS SEMINAR
Exploring Responsibility and Complicity
Brussels, Belgium

09.00 Registrations
Tea & Coffee on arrival

09.15 Welcome – Robyn Van Maryns

09.30 **Keynote speeches: Setting the context for business responsibility with regards to complicity in human rights abuses**
Moderator: Mary Robins (Honorary Chair, Business Leaders Initiative on Human Rights & The 2005 Business and Human Rights Seminar)
- Wilma Wallace (Associate General Counsel, Vice President for Legal, Government & Environmental Affairs, Gap Inc.)
- Luis Moreno-Ocampo (Chief Prosecutor, International Criminal Court)
- Paul Watchman (Partner, Freshfields Bruckhaus Deringer)
Participant questions and discussion

10.45 Tea & Coffee

11.15 **Panel: The various interpretations of legal and moral complicity in a business context**
Moderator: Robin Van Maryns (Director, TwentyFifty Ltd)
- Luis Moreno-Ocampo (Chief Prosecutor, International Criminal Court)
- Steven Jin (Chief Legal Counsel, Amnesty International)
Full audience participation and discussion

12.45 Lunch

14.00 **Keynote speeches: Examining complicity and responsibility dilemmas for business**
Moderator: Robin Van Maryns (Director, TwentyFifty Ltd)
- Peter Sutherland KCMG (Chairman of BP plc & Chairman of Goldman Sachs International)
- Jorge Daniel Taillant (Executive Director, Center for Human Rights and Environment, Argentina)
Roundtable discussion and feedback on three business dilemmas
On-stage responses to the dilemmas from:
- Lise Kingo (Executive Vice President, People, Reputation & Relations, Novo Nordisk A/S)
- Alan Detheridge (Vice President, External Affairs, Exploration and Production, Shell International)
Full audience discussion

15.40 Tea & Coffee

15.50 **The Way Forward - The Mandate of the Special Representative and the Company Reports**
Moderator: Robin Van Maryns (Director, TwentyFifty Ltd)
- Steven Jin (Chief Legal Counsel, Amnesty International) - Company Report
- Professor John Ruggie (Special Representative on the Issue of Human Rights, Transnational Corporations and Other Business Enterprises)
- Steven Jin (Chief Legal Counsel, Amnesty International) - Company Report
Full room discussion

16.15 Closing remarks by Mary Robinson

16.30 Seminar ends

Annex 5: Manipulated sentences in text Amnesty

The list of all manipulated sentences in the text Amnesty, together with their translation into Czech and Dutch for matching purposes.

Syntactic manipulations

The evidence which led to Shi's conviction and which was provided by Yahoo! Hong Kong, included details of his Yahoo mail account and the IP address, despite a confidentiality agreement between Shi and Yahoo!.

Důkazy, které vedly k Shiově odsouzení a které dodal Yahoo! Hong Kong, obsahovaly údaje o jeho emailovém účtu na Yahoo! i identifikaci počítače, a to navzdory smlouvě o utajení mezi Shim a Yahoo!.

Het bewijs dat tot de veroordeling van Shi leidde en dat aangeleverd werd door Yahoo! Hong Kong, bevatte details van zijn Yahoo mailaccount en het IP adres, ondanks het geheimhoudingsakkoord tussen Shi en Yahoo!

Nigerian soldiers, who served in special military forces and who were heavily armed with automatic rifles, used live ammunition against protesters from a small ethnic community for entering Chevron's oil terminal.

Nigerijští vojáci, kteří slouží ve speciálních ozbrojených silách a kteří byli těžce ozbrojeni lehkými kulomety, použili ostré náboje proti protestantům z malé etnické komunity, protože vstoupili do prostoru ropného těžebného zařízení Chevronu.

Nigeriaanse soldaten die deel uitmaakten van een speciale eenheid en die zwaar bewapend waren met automatische geweren, gebruikten scherpe munitie tegen betogers van een kleine etnische gemeenschap die Chevron's olieterminal waren binnengedrongen.

Non-state actors which include privately-owned companies and which are considered an organ of society, must uphold human rights within their area of influence in the context of both their operations and local communities.

Nestátní entity, mezi které patří i soukromé společnosti a které jsou považovány za orgány společnosti, musí dodržovat lidská práva ve sféře svého vlivu v kontextu své podnikatelské činnosti i v rámci místní komunity.

Niet-gouvernementele actoren waaronder ook bedrijven vallen en die een belangrijke rol spelen in de maatschappij, moeten mensenrechten respecteren in zowel hun handelingen als in hun contacten met de buurt waarin ze gevestigd zijn.

Corporate complicity, which is now attracting attention of the international community and which is an emerging area of law, poses important moral and legal questions about borderlines between right and wrong.

Firemní spoluvina, která nyní přitahuje pozornost mezinárodního společenství a která je nově vznikající oblastí práva, klade důležité morální i právní otázky o hranici mezi tím, co je správné a co ne.

Deze medeplichtigheid van bedrijven, die nu aandacht krijgt van de internationale gemeenschap en die steeds belangrijker wordt in recht, adresseert belangrijke morele en wettelijke vragen over de grens tussen goed en fout.

Values and principles which often reach beyond existing law and which are worth fighting for should be the basis of our efforts for improved legislation in an international context.

Hodnoty a principy, které často přesahují platné právo a za něž stojí za to bojovat, musí být základem našeho úsilí o lepší právní úpravu v mezinárodním měřítku.

Waarden en principes die vaak buiten de wetgeving vallen en die het waard zijn voor te vechten, zouden de basis moeten zijn voor onze inzet voor betere wetgeving in een internationale context.

Companies which use migrant labour directly or which rely on suppliers of workers, must check the identity of their workers very carefully because of an increase in human trafficking.

Podniky, které najímají migrující pracovníky přímo nebo které využívají dodavatelů pracovní síly, musí pečlivě kontrolovat totožnost těchto pracovníků z důvodu nárůstu obchodu s lidmi.

Bedrijven die rechtstreeks gebruik maken van migrantenarbeid of die via tussenpersonen werken, moeten de identiteit van hun arbeiders zorgvuldig controleren omwille van toenemende gevallen van mensensmokkel.

Many companies, which had operations in South Africa during the apartheid and which followed the laws of that time, effectively aided the South African government in implementing its discriminatory policies.

Mnoho společností, které podnikaly v Jižní Africe v době apartheidu a které se držely soudobých zákonů, fakticky napomáhalo jihoafrické vládě v prosazování její diskriminační politiky.

Veel bedrijven, die werkzaam waren in Zuid-Afrika tijdens de Apartheid en die de wetten van toen volgden, hielpen de Zuid-Afrikaanse regering daadwerkelijk bij het implementeren van haar discriminerende beleid.

Emerging jurisprudence in criminal law, which builds on complicity for crimes against humanity and which has provided a broad definition of complicity, helps develop civil standards and principles in different tribunals and courts.

Rozvíjející se právní věda v oblasti trestního práva, která staví na spoluvině u zločinů proti lidskosti a která je zdrojem široké definice spoluviny, pomáhá rozvíjet občanské standardy a principy u soudů různých stupňů.

De opkomende jurisprudentie in het strafrecht, die steunt op medeplichtigheid aan daden tegen de menselijkheid en die een brede definitie van medeplichtigheid heeft opgesteld, helpt normen en principes te ontwikkelen in verschillende rechtbanken.

A company which wishes to build a golf course and which accepts that the state evicts people from that land, derives a benefit from an abuse of human rights.

Společnost, která hodlá vybudovat golfové hřiště a která připustí, aby stát z tohoto pozemku vysídlil obyvatele, má z porušování lidských práv výhody.

Een bedrijf dat een golfbaan wil aanleggen en aanvaardt dat de staat mensen van dat grondgebied verjaagt, haalt voordeel uit de schending van een mensenrecht.

The Kimberley process, in which companies have joined with governments and NGOs and which established voluntary codes of conduct for the diamond industry, helps avoid complicity in human rights abuses from diamond trade in conflict ridden countries.

Kimberleyský proces, kde se spojily firmy s vládami a nevládními organizacemi a který vytvořil dobrovolný kodex pro diamantový průmysl, pomáhá předcházet spoluvině při porušování lidských práv formou obchodu s diamanty ve válkou zmítaných zemích.

Het Kimberley Proces waarin bedrijven, overheden en NGO’s samenwerken en vrijwillige gedragscodes voor de diamantsector opgesteld hebben, helpt medeplichtigheid aan schending van de mensenrechten in de diamanthandel in conflictgebieden te voorkomen.

Lexical manipulations (figures)

At least 30 others were injured, 15 of them seriously, by blows from rifle butts and other weapons.

Nejméně 30 dalších bylo zraněno, z toho 15 vážně, ranami pažbou samopalů a jiných zbraní.
Minstens 30 anderen waren gekwetst, waarvan 15 ernstig, door geweerkogels en andere wapens.

The boundary fence was cut in 35 places, and 87 windows and 17 helicopter windscreens were smashed.
Vnější plot byl na 35 místech prostřižen, 87 oken a 17 čelních skel helikoptér bylo rozbito.
De afrastering was op 35 plaatsen doorgesneden en 87 vensters en 17 helikopterruiten werden ingeslagen.

The protesters said that Chevron Nigeria had not provided the 1750 jobs and 55 development projects they were promised.
Protestanti prohlásili, že Chevron Nigérie nevytvořil 1750 pracovních míst ani 55 rozvojových projektů, které jim byly přislíbeny.
De demonstranten zeiden dat Chevron Nigeria de 1750 banen en 55 ontwikkelingsprojecten die het beloofd had nog niet gerealiseerd had.

Some analysts estimate that 13 to 15 of every 100 companies present in zones of conflict may indirectly contribute to armed conflicts.
Podle některých analytiků 13 až 15 z každé stovky firem přítomných ve válečných zónách potenciálně přispívá na ozbrojený konflikt.
Een aantal analisten schatten dat 13 tot 15 % van de bedrijven die gevestigd zijn in conflictzones mogelijk indirect bijdragen tot gewapende conflicten.

The right not to be held in slavery or servitude was recognized in Article 2 of the 1926 Slavery Convention.
Právo nebýt držen v otroctví či vazalství bylo uznáno v článku 2 konvence proti otroctví z roku 1926.
Het recht om niet in slavernij of onderworpenheid te leven werd erkend in Artikel 2 van de slavernij conventie van 1926.

Statistics show that women are 27% less likely, and elderly persons 63% less likely to be hired.
Statistiky ukazují, že ženy mají o 27 % menší šanci a staré osoby o 63 % menší šanci získat zaměstnání.
Statistieken tonen aan dat vrouwen 27 % en ouderen 63 % minder kans hebben om aangeworven te worden.

The company saw its turnover drop by 54% and suffered a USD 127 million loss in just 2 years.
Společnost zaznamenala pokles obratu o 54 % a utrpěla ztrátu 127 miliónů dolarů za pouhé dva roky.
Het bedrijf zag zijn omzet dalen met 54 % en leed een verlies van 127.000.000 USD in slechts 2 jaar.

The growing body of research now includes 138 academic studies and at least 23 court findings.
Narůstající objem vědeckých prací už zahrnuje 138 akademických studií a nejméně 23 soudních nálezů.
De toenemende hoeveelheid research omvat nu 138 academische studies en minstens 23 rechtbankverslagen.

Nike, the sportswear manufacturer, has 657 suppliers spread across 65 countries, not including the US, where it is headquartered.
Výrobce sportovního oblečení Nike má 657 dodavatelů v 65 zemích bez započítání USA, kde firma sídlí.

Nike, de producent van sportkledij, heeft 657 leveranciers verspreid over 65 landen, exclusief de VS waar het zijn hoofdkwartier heeft.

Some estimates suggest that 46% of companies operating in Asia, 51% in South America and 63% in Africa may be in this position.

Podle některých odhadů je 46 % firem působících v Asii, 51 % v Jižní Americe a 63 % v Africe potenciálně v této pozici.

Sommige schattingen suggereren dat 46 % van de bedrijven die werkzaam zijn in Azië, 51 % in Zuid-Amerika en 63 % in Afrika zich in deze positie bevinden.

Semantic manipulations

The company did not promise not to assist the security forces in the future.

Firma se nezavázala, že ozbrojeným silám nebude v budoucnu napomáhat.

Het bedrijf beloofde niet de veiligheidsdiensten in de toekomst niet te helpen.

Or the company does not disaffiliate itself from such governments or armed groups?

Či co když se firma nedistancuje od takové vlády či ozbrojené skupiny?

Of het bedrijf distantieert zich niet van zulke overheden of gewapende groeperingen?

On the other, companies simply did not disapprove of abuse committed by others.

V druhém extrému firma pouze neodsoudila porušování práv, kterého se dopustili jiní.

Anderzijds keurden bedrijven misbruik door anderen gewoon niet af.

But many countries do not undertake not to apply the penalty.

Ale mnoho zemí se nezavázalo, že trest smrti nebude uplatňovat.

Maar veel landen verbinden zich niet tot het niet-toepassen van de straf.

Yet, many companies do not disapprove of a number of practices.

Avšak mnoho firem neodsuzuje celou řadu postupů.

Toch keuren veel bedrijven een aantal praktijken niet af.

Many companies in the developing world do not respect the rule not to employ children.

Mnoho společností v rozvojových zemích nerespektuje pravidlo nezaměstnávat děti.

Veel bedrijven in ontwikkelingslanden respecteren de regel niet om geen kinderen in dienst te nemen.

A complicit company does not explicitly disapprove of such child labour.

Spoluviná firma se jednoznačně nestaví proti takovému zaměstnávání dětí.

Een medeplichtig bedrijf keurt dergelijke kinderarbeid niet expliciet af.

They have not undertaken not to assist the Chinese governments in restricting freedom of expression.

Nezavázali se, že nebudou čínské vládě napomáhat při omezování svobody projevu.

Ze hebben zich er niet toe verbonden de Chinese overheden niet te steunen in het beperken van de vrije meningsuiting.

Sometimes, a company simply does not disapprove of a certain practice.

Někdy stačí, že firma prostě neodsoudí některé postupy.

Soms keurt een bedrijf bepaalde praktijken gewoon niet af.

Importantly, Nike had not explicitly promised not to allow this happen again.

Důležité bylo to, že Nike jednoznačně neslíbil, že nepřipustí, aby se situace opakovala.

Let wel, Nike had niet expliciet beloofd dit niet meer te laten gebeuren.

Annex 6: Manipulated lists in companies texts

English	Dutch	Czech
Text China		
right to a safe work environment	recht op een veilige werkomgeving	právo na bezpečné pracovní prostředí
non-discrimination in employment	non-discriminatie op het werk	nediskriminace v zaměstnání
right to physical and mental health	recht op lichamelijke en geestelijke gezondheid	právo na tělesné a duševní zdraví
abolition of forced labour	afschaffing van dwangarbeid	odstranění nucené práce
right to collective barganing	recht op collectieve onderhandeling	právo na kolektivní vyjednávání
right to development	recht op ontwikkeling	právo na rozvoj
right to equal pay for equal work	recht op gelijk loon voor gelijk werk	právo na stejný plat za stejnou práci
right to political life	recht op het politieke leven	právo na politický život
Text Brazil		
right to organise	vakverenigingrecht	právo organizovat se
equality at work	gelijkheid op het werk	rovnost v práci
right to work and vocational training	recht op werk en op een beroepsopleiding	právo na práci a odbornou přípravu
abolition of child labour	afschaffing van kinderarbeid	odstranění dětské práce
prohibition of degrading treatment	verbod van onerende behandeling	zákaz ponižujícího zacházení
right to education	recht op onderwijs	právo na vzdělání
freedom of association	vrijheid van vereniging	svoboda sdružování
right to equal protection under the law	recht op gelijke bescherming onder de wet	právo na rovnou ochranu podle zákona
Text China		
extractive industry	mijnindustrie	těžební průmysl
financial services	financiële diensten	finanční služby
computer hardware	computer hardware	počítačový hardware
transport and logistics	transport en logistiek	doprava a logistika
infrastructure and utilities	infrastructuur en nutsvoorziening	infrastruktura a veřejné služby
it and telecommunications	it en telecommunicatie	it a telekomunikace
pharmaceutical industry	farmaceutische industrie	farmaceutický průmysl
chemical industry	chemische industrie	chemický průmysl
Text Brazil		
automotive industry	auto-industrie	automobilový průmysl
insurance services	verzekeringsdiensten	pojišťovací služby
forestry and fishing	bosbouw en visserij	lesnictví a rybářství
food and beverage industry	voedsel- en drankenindustrie	potravinářský a nápojový průmysl
media and advertising	media en reclame	média a reklama
construction industry	bouw	stavebnictví
wholesale and retail	groot- en kleinhandel	velkoobchod a maloobchod
legal services	juridische dienstverlening	právní služby

Annex 7: Introductory letter

Dear colleague,

My name is Šárka Timarová and I am writing to you to ask whether you would kindly take part in a research study of interpreting that I am conducting. The study is part of my doctoral research project and will be submitted as a PhD dissertation to Charles University, Prague, Czech Republic. The aim of the study is to investigate cognitive processes involved in simultaneous interpreting. More specifically, you will be asked to do two things. First of all, you will be asked to simultaneously interpret several texts, from English into your mother tongue. Secondly, you will be asked to perform several tasks measuring various cognitive processes. Your performance on the two types of tasks (interpreting and cognitive processes) will then be compared by means of statistical analyses, which would allow me to see whether the measured cognitive processes contribute to the process of simultaneous interpreting.
At this point, I would like to stress three important aspects of the study. For one, unfortunately, there is no research budget for the study, and I will not be able to pay you for your participation. Secondly, the study is a fairly extensive one, and it would require that you allow me to test you **in two sessions, each lasting approximately 90 minutes**. And thirdly, I would like to make it clear from the beginning that this is not a study of quality. The aim is **not** to look for and point out errors that interpreters make. You are being asked to take part because you are a professional interpreter who has all the necessary qualifications to perform the job. My aim is to understand better what we do when we interpret. Understanding the process of interpreting is important for developing better training practices and solid evidence is also helpful when arguing eg for certain working conditions.
I would be very grateful if you could find the time to take part in this study. I will give you a few days to think about it, and will contact you again in 7 days, unless I hear from you before that. I will be happy to answer any questions you may have. You can reach me at this email address, or by telephone at […].

Best regards,
Sarka Timarova

Annex 8: Consent form

CONSENT FORM

WORKING MEMORY AND SIMULTANEOUS INTERPRETING

I have agreed to participate, as a volunteer and without reimbursement, in this study on working memory in simultaneous interpreting. The study is conducted by Šárka Timarová, and will be submitted as a PhD thesis to the Charles University, Prague, Czech Republic.
In addition to the above stated purpose, I agree that the data collected can be used in the following manner:

other research projects	yes/no
including/excluding research into quality	
the data can be made available to other researchers	yes/no
voice samples can be used for dissemination of results	yes/no

I have been informed about the research aims of the study. I have been also informed that I will be asked to perform several laboratory-based tasks aimed at a) providing a controlled sample of my simultaneous interpreting performance, b) measuring various aspects of my working memory. The nature of the working memory tasks will almost inevitably lead to my not being able to complete the tasks without errors. I am aware that this is necessary for measurement purposes, and that my inability to complete all tasks to perfection does not allow any judgement about me as a person, my intelligence or cognitive abilities. It is also understood that my simultaneous interpreting performance may differ, for a variety of reasons, from my interpreting performance in a real job situation.
I have received all information necessary to complete the tasks. Any information that will be collected about my person, including my personal details and my scores on the tasks, will be considered confidential and will not be identified with me. I am aware that I may withdraw from the study at any point, without giving any reason, and that in such a case all data collected about my person up to that point will be immediately destroyed. I am also aware of my right to ask any questions and receive answers, before and after I have participated in the study (with the exception of questions which might jeopardise the aims of the study). My questions can be directed to the author of the study, Ms Šárka Timarová, in person or by e-mail at [*author's email address*], and any complaints may be directed to the author's supervisor, Prof. Ivana Čeňková, by e-mail at [*supervisor's email address*].
I sign this document below to indicate that I have understood the conditions and that I agree to participate in the study as described in this document. I will receive a copy of this document.

Date:	Place:
Participant Signature	Experimenter Signature
Participant Name	Šárka Timarová

Annex 9: Demographic information sheet

PARTICIPANT ID: ________

ABOUT YOU

Age: ___________________ Sex: M F

ABOUT YOUR EDUCATION

Highest level of completed education: secondary
undergraduate (bachelor)
postgraduate (masters)
doctoral

Subject area for which the highest degree was awarded
____

Subject area for which the second highest degree was awarded
____

Are you a trained simultanous interpreter? Yes No

If yes, where and when did you train?
____

Length of the training programme:
____

ABOUT YOUR ENGLISH

I comprehend spoken English (neutral accent):
1(poor) 2 3 4 5 6 7 8 9 10(like a native)

Where and how did you learn English? age ____

school____

immersion___________________________________

I interpret from English every time I work. yes/no

Estimate % of working time you interpret from English.
____

Is English your preferred relay language? yes/no

ABOUT YOUR WORKING LANGUAGES

What is your mother tongue? ____

What are your working languages in order of their strength (start with the strongest)? Are they your A, B or C languages? Were they in your training combination or did you include them later?

Language	A/B/C?	Trained?
		Yes No
		Yes No
		Yes No
		Yes No
		Yes No

ABOUT YOUR PROFESSIONAL EXPERIENCE AS A SIMULTANEOUS INTERPRETER

How many years have you been interpreting? ________________________________

Can you estimate how many days you work per year?

Example:
2007 – 100 days
2004-2006 – 60 days per year

Period	Days per year

Annex 10: English vocabulary test

ANSWER SHEET

VOCABULARY TEST

In this test, you will see a list of "words", some of which are real and some of which are invented. All the "words" are verbs. Place a ✘ next to the **INVENTED** words.

You will have **3 minutes** to complete the test.

1.	to campaign	_____	26.	to settle	_____	51.	to digame	_____
2.	to futt	_____	27.	to driggle	_____	52.	to numbelate	_____
3.	to bourble	_____	28.	to witness	_____	53.	to colour	_____
4.	to fear	_____	29.	to emerge	_____	54.	to wordle	_____
5.	to preyout	_____	30.	to prinkle	_____	55.	to complement	_____
6.	to study	_____	31.	to pronate	_____	56.	to repair	_____
7.	to savedown	_____	32.	to complicate	_____	57.	to reform	_____
8.	to compile	_____	33.	to squeeze	_____	58.	to quote	_____
9.	to motivate	_____	34.	to congratulate	_____	59.	to address	_____
10.	to decite	_____	35.	to keepsick	_____	60.	to waste	_____
11.	to megalize	_____	36.	to hesitate	_____	61.	to announce	_____
12.	to markle	_____	37.	to chariover	_____	62.	to mayto	_____
13.	to abolish	_____	38.	to strang	_____	63.	to type	_____
14.	to root	_____	39.	to permit	_____	64.	to wait	_____
15.	to distinguish	_____	40.	to oldenate	_____	65.	to eaude	_____
16.	to outlate	_____	41.	to skey	_____	66.	to kinnear	_____
17.	to sink	_____	42.	to unleash	_____	67.	to stay	_____
18.	to encompass	_____	43.	to honch	_____	68.	to monadate	_____
19.	to review	_____	44.	to name	_____	69.	to box	_____
20.	to celebrate	_____	45.	to organize	_____	70.	to authorise	_____
21.	to demolish	_____	46.	to mention	_____	71.	to commission	_____
22.	to administer	_____	47.	to struggle	_____	72.	to trace	_____
23.	to erode	_____	48.	to yell	_____	73.	to judge	_____
24.	to fabulation	_____	49.	to promise	_____	74.	to conceive	_____
25.	to join	_____	50.	to violate	_____	75.	to inherit	_____

Annex 11: Letter span task answer sheets

ANSWER SHEET

LETTER SPAN

Example:

You see: ABCDEF. You remember: AB something DFE. Write:

1. __A___ ___B__ _______ __D___ __F___ __E___ _______ _______ _______ _______

Leave a blank space if you cannot remember the letter. It is important to remember the right letter at the right position!

1. ______ ______ ______ ______ ______ ______ ______ ______ ______ ______

2. ______ ______ ______ ______ ______ ______ ______ ______ ______ ______

1. ______ ______ ______ ______ ______ ______ ______ ______ ______ ______

2. ______ ______ ______ ______ ______ ______ ______ ______ ______ ______

3. ______ ______ ______ ______ ______ ______ ______ ______ ______ ______

4. ______ ______ ______ ______ ______ ______ ______ ______ ______ ______

5. ______ ______ ______ ______ ______ ______ ______ ______ ______ ______

6. ______ ______ ______ ______ ______ ______ ______ ______ ______ ______

7. ______ ______ ______ ______ ______ ______ ______ ______ ______ ______

8. ______ ______ ______ ______ ______ ______ ______ ______ ______ ______

9. ______ ______ ______ ______ ______ ______ ______ ______ ______ ______

10. ______ ______ ______ ______ ______ ______ ______ ______ ______ ______

Annex 13: Statistical analysis: working memory

Spearman correlation matrix for working memory tasks.

		Age	Experience Years	Experience Days	EN test	Arrow	Cattell	Letter	Corsi	Complex	Antisacc	ArrowFla	Barrouillet	Nback	Number-letter
Age	Spearman	1.000	.833**	.695**	.193	.621**	-.573**	-.260	-.378*	-.336	-.427*	-.318	-.121	-.128	.287
	Sig. (2-tailed)	.	.000	.000	.325	.000	.001	.181	.047	.080	.026	.106	.555	.524	.146
	N	28	28	28	28	28	28	28	28	28	27	27	26	27	27
Experience Years	Spearman	.833**	1.000	.893**	.339	.399*	-.575**	-.198	-.300	-.372	-.377	-.503**	.117	-.182	.138
	Sig. (2-tailed)	.000	.	.000	.078	.035	.001	.313	.121	.052	.053	.008	.570	.363	.493
	N	28	28	28	28	28	28	28	28	28	27	27	26	27	27
Experience Days	Spearman	.695**	.893**	1.000	.343	.337	-.466*	-.219	-.252	-.381*	-.336	-.602**	.101	-.055	-.027
	Sig. (2-tailed)	.000	.000	.	.074	.080	.012	.262	.196	.045	.087	.001	.622	.787	.892
	N	28	28	28	28	28	28	28	28	28	27	27	26	27	27
EN test	Spearman	.193	.339	.343	1.000	-.265	-.037	.100	.535**	.110	.017	-.264	-.078	-.025	-.166
	Sig. (2-tailed)	.325	.078	.074	.	.173	.853	.614	.003	.578	.933	.184	.706	.900	.407
	N	28	28	28	28	28	28	28	28	28	27	27	26	27	27
Arrow	Spearman	.621**	.399*	.337	-.265	1.000	-.372	-.149	-.655**	-.277	-.481*	-.138	-.063	-.203	.302
	Sig. (2-tailed)	.000	.035	.080	.173	.	.051	.448	.000	.153	.011	.491	.759	.309	.126
	N	28	28	28	28	28	28	28	28	28	27	27	26	27	27
Cattell	Spearman	-.573**	-.575**	-.466*	-.037	-.372	1.000	.275	.316	.357	.267	-.029	.038	.330	.046
	Sig. (2-tailed)	.001	.001	.012	.853	.051	.	.157	.102	.062	.178	.887	.856	.092	.820
	N	28	28	28	28	28	28	28	28	28	27	27	26	27	27
Letter	Spearman	-.260	-.198	-.219	.100	-.149	.275	1.000	.133	.686**	.063	-.211	-.056	.185	-.180
	Sig. (2-tailed)	.181	.313	.262	.614	.448	.157	.	.499	.000	.757	.292	.787	.357	.370
	N	28	28	28	28	28	28	28	28	28	27	27	26	27	27
Corsi	Spearman	-.378*	-.300	-.252	.535**	-.655**	.316	.133	1.000	.147	.305	.022	-.119	.006	-.203
	Sig. (2-tailed)	.047	.121	.196	.003	.000	.102	.499	.	.454	.122	.912	.564	.978	.309
	N	28	28	28	28	28	28	28	28	28	27	27	26	27	27

		Age	Experience Years	Experience Days	EN test	Arrow	Cattell	Letter	Corsi	Complex	Antisacc	ArrowFla	Barrouillet	Nback	Number-letter
Complex	Spearman	-.336	-.372	-.381*	.110	-.277	.357	.686**	.147	1.000	.234	.128	-.226	.283	.103
	Sig. (2-tailed)	.080	.052	.045	.578	.153	.062	.000	.454	.	.240	.525	.266	.153	.609
	N	28	28	28	28	28	28	28	28	28	27	27	26	27	27
Antisacc	Spearman	-.427*	-.377	-.336	.017	-.481*	.267	.063	.305	.234	1.000	.214	.058	.057	-.109
	Sig. (2-tailed)	.026	.053	.087	.933	.011	.178	.757	.122	.240	.	.293	.783	.782	.595
	N	27	27	27	27	27	27	27	27	27	27	26	25	26	26
ArrowFla	Spearman	-.318	-.503**	-.602**	-.264	-.138	-.029	-.211	.022	.128	.214	1.000	-.069	-.030	.056
	Sig. (2-tailed)	.106	.008	.001	.184	.491	.887	.292	.912	.525	.293	.	.744	.885	.784
	N	27	27	27	27	27	27	27	27	27	26	27	25	26	26
Barrouillet	Spearman	-.121	.117	.101	-.078	-.063	.038	-.056	-.119	-.226	.058	-.069	1.000	.098	-.106
	Sig. (2-tailed)	.555	.570	.622	.706	.759	.856	.787	.564	.266	.783	.744	.	.635	.613
	N	26	26	26	26	26	26	26	26	26	25	25	26	26	25
Nback	Spearman	-.128	-.182	-.055	-.025	-.203	.330	.185	.006	.283	.057	-.030	.098	1.000	-.372
	Sig. (2-tailed)	.524	.363	.787	.900	.309	.092	.357	.978	.153	.782	.885	.635	.	.061
	N	27	27	27	27	27	27	27	27	27	26	26	26	27	26
Number-letter	Spearman	.287	.138	-.027	-.166	.302	.046	-.180	-.203	.103	-.109	.056	-.106	-.372	1.000
	Sig. (2-tailed)	.146	.493	.892	.407	.126	.820	.370	.309	.609	.595	.784	.613	.061	.
	N	27	27	27	27	27	27	27	27	27	26	26	25	26	27

Annex 14: Statistical analysis: simultaneous interpreting

Spearman correlation matrix for simultaneous interpreting tasks.

		Age	Experience Years	Experience Days	EN test	Arrow	Cattell	Syntax	Figures	Negatives	Vocabulary Ratio	Vocabulary Unique	Comp Difference	Comp Average	Median EVS	SD EVS
Age	Spearman	1.000	.833**	.695**	.193	.621**	-.573**	.397*	.073	.396*	.349	.344	-.376	.385	-.216	-.196
	Sig. (2-tailed)	.	.000	.000	.325	.000	.001	.037	.713	.037	.132	.138	.058	.052	.269	.317
	N	28	28	28	28	28	28	28	28	28	20	20	26	26	28	28
Experience Years	Spearman	.833**	1.000	.893**	.339	.399*	-.575**	.438*	.287	.587**	.400	.427	-.324	.524**	-.360	-.414*
	Sig. (2-tailed)	.000	.	.000	.078	.035	.001	.020	.138	.001	.081	.060	.107	.006	.060	.029
	N	28	28	28	28	28	28	28	28	28	20	20	26	26	28	28
Experience Days	Spearman	.695**	.893**	1.000	.343	.337	-.466*	.372	.282	.624**	.256	.269	-.385	.510**	-.434*	-.412*
	Sig. (2-tailed)	.000	.000	.	.074	.080	.012	.051	.146	.000	.276	.251	.052	.008	.021	.029
	N	28	28	28	28	28	28	28	28	28	20	20	26	26	28	28
EN test	Spearman	.193	.339	.343	1.000	-.265	-.037	.159	.171	.240	.056	-.025	-.391*	.336	-.382*	-.524**
	Sig. (2-tailed)	.325	.078	.074	.	.173	.853	.420	.384	.218	.815	.918	.048	.094	.045	.004
	N	28	28	28	28	28	28	28	28	28	20	20	26	26	28	28
Arrow	Spearman	.621**	.399*	.337	-.265	1.000	-.372	.305	-.032	.031	-.005	.305	-.074	.189	.101	.207
	Sig. (2-tailed)	.000	.035	.080	.173	.	.051	.114	.872	.875	.982	.191	.719	.356	.611	.291
	N	28	28	28	28	28	28	28	28	28	20	20	26	26	28	28
Cattell	Spearman	-.573**	-.575**	-.466*	-.037	-.372	1.000	-.234	-.143	-.204	-.332	-.286	.185	-.050	.184	.153
	Sig. (2-tailed)	.001	.001	.012	.853	.051	.	.231	.468	.299	.153	.222	.366	.807	.347	.438
	N	28	28	28	28	28	28	28	28	28	20	20	26	26	28	28
Syntax	Spearman	.397*	.438*	.372	.159	.305	-.234	1.000	.248	.338	.440	.337	.078	.626**	-.568**	-.230
	Sig. (2-tailed)	.037	.020	.051	.420	.114	.231	.	.203	.079	.052	.146	.704	.001	.002	.239
	N	28	28	28	28	28	28	28	28	28	20	20	26	26	28	28

		Age	Experience Years	Experience Days	EN test	Arrow	Cattell	Syntax	Figures	Negatives	Vocabulary Ratio	Vocabulary Unique	Comp Difference	Comp Average	Median EVS	SD EVS
Figures	Spearman	.073	.287	.282	.171	-.032	-.143	.248	1.000	.514**	-.194	.266	-.229	.686**	-.515**	-.438*
	Sig. (2-tailed)	.713	.138	.146	.384	.872	.468	.203	.	.005	.412	.258	.261	.000	.005	.020
	N	28	28	28	28	28	28	28	28	28	20	20	26	26	28	28
Negatives	Spearman	.396*	.587**	.624**	.240	.031	-.204	.338	.514**	1.000	.359	.341	-.074	.537**	-.498**	-.337
	Sig. (2-tailed)	.037	.001	.000	.218	.875	.299	.079	.005	.	.120	.141	.720	.005	.007	.079
	N	28	28	28	28	28	28	28	28	28	20	20	26	26	28	28
Vocabulary Ratio	Spearman	.349	.400	.256	.056	-.005	-.332	.440	-.194	.359	1.000	.287	.214	-.019	-.243	-.215
	Sig. (2-tailed)	.132	.081	.276	.815	.982	.153	.052	.412	.120	.	.220	.379	.937	.303	.364
	N	20	20	20	20	20	20	20	20	20	20	20	19	19	20	20
Vocabulary Unique	Spearman	.344	.427	.269	-.025	.305	-.286	.337	.266	.341	.287	1.000	-.025	.291	-.013	-.100
	Sig. (2-tailed)	.138	.060	.251	.918	.191	.222	.146	.258	.141	.220	.	.919	.226	.957	.674
	N	20	20	20	20	20	20	20	20	20	20	20	19	19	20	20
Comp Difference	Spearman	-.376	-.324	-.385	-.391*	-.074	.185	.078	-.229	-.074	.214	-.025	1.000	-.363	.154	.260
	Sig. (2-tailed)	.058	.107	.052	.048	.719	.366	.704	.261	.720	.379	.919	.	.068	.452	.199
	N	26	26	26	26	26	26	26	26	26	19	19	26	26	26	26
Comp Average	Spearman	.385	.524**	.510**	.336	.189	-.050	.626**	.686**	.537**	-.019	.291	-.363	1.000	-.567**	-.378
	Sig. (2-tailed)	.052	.006	.008	.094	.356	.807	.001	.000	.005	.937	.226	.068	.	.003	.057
	N	26	26	26	26	26	26	26	26	26	19	19	26	26	26	26
Median EVS	Spearman	-.216	-.360	-.434*	-.382*	.101	.184	-.568**	-.515**	-.498**	-.243	-.013	.154	-.567**	1.000	.651**
	Sig. (2-tailed)	.269	.060	.021	.045	.611	.347	.002	.005	.007	.303	.957	.452	.003	.	.000
	N	28	28	28	28	28	28	28	28	28	20	20	26	26	28	28
SD EVS	Spearman	-.196	-.414*	-.412*	-.524**	.207	.153	-.230	-.438*	-.337	-.215	-.100	.260	-.378	.651**	1.000
	Sig. (2-tailed)	.317	.029	.029	.004	.291	.438	.239	.020	.079	.364	.674	.199	.057	.000	.
	N	28	28	28	28	28	28	28	28	28	20	20	26	26	28	28

Annex 15: Statistical analysis: working memory and simultaneous interpreting

Spearman correlation matrix for working memory vs simultaneous interpreting tasks.

		Letter	Corsi	Complex	Antisacc	ArrowFla	Barrouillet	Nback	Number-letter
Syntax	Spearman	.193	-.120	-.190	-.170	-.301	.055	-.039	-.124
	Sig. (2-tailed)	.325	.542	.334	.395	.127	.788	.849	.538
	N	28	28	28	27	27	26	27	27
Figures	Spearman	.337	-.184	.183	-.062	-.092	.275	.522**	-.409*
	Sig. (2-tailed)	.080	.350	.352	.759	.649	.174	.005	.034
	N	28	28	28	27	27	26	27	27
Negatives	Spearman	.026	-.108	-.114	.088	-.304	.372	.093	-.024
	Sig. (2-tailed)	.897	.583	.563	.663	.123	.061	.643	.906
	N	28	28	28	27	27	26	27	27
Vocabulary Ratio	Spearman	-.301	-.103	-.370	-.091	-.142	.193	-.648**	.201
	Sig. (2-tailed)	.198	.665	.108	.711	.562	.444	.003	.409
	N	20	20	20	19	19	18	19	19
Vocabulary Unique	Spearman	-.180	-.368	-.284	-.173	.013	.334	-.109	.115
	Sig. (2-tailed)	.447	.110	.225	.479	.957	.176	.657	.638
	N	20	20	20	19	19	18	19	19
Comp Difference	Spearman	.106	-.068	.006	.167	.137	.543**	-.126	.205
	Sig. (2-tailed)	.606	.740	.975	.424	.514	.005	.550	.326
	N	26	26	26	25	25	25	25	25
Comp Average	Spearman	.372	-.223	.061	.003	-.423*	.099	.312	-.216
	Sig. (2-tailed)	.061	.273	.768	.990	.035	.637	.129	.300
	N	26	26	26	25	25	25	25	25
Median EVS	Spearman	-.013	-.031	.071	.071	.092	-.227	-.266	.415*
	Sig. (2-tailed)	.948	.876	.719	.724	.647	.266	.181	.031
	N	28	28	28	27	27	26	27	27
SD EVS	Spearman	-.068	-.101	-.009	.141	.156	-.055	-.152	.116
	Sig. (2-tailed)	.730	.607	.962	.483	.438	.790	.449	.565
	N	28	28	28	27	27	26	27	27

Regression: companies average*arrow flanker, median EVS

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	Rank of ArrowFla[a]	.	Enter
2	Rank of MedianEVS[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: Rank of Companies Average

Model Summary

					Change Statistics				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change
1	.417[a]	.174	.138	7.095311	.174	4.850	1	23	.038
2	.664[b]	.441	.390	5.970656	.266	10.481	1	22	.004

a. Predictors: (Constant), Rank of ArrowFla

b. Predictors: (Constant), Rank of ArrowFla, Rank of MedianEVS

ANOVA[c]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	244.181	1	244.181	4.850	.038[a]
	Residual	1157.899	23	50.343		
	Total	1402.080	24			
2	Regression	617.808	2	308.904	8.665	.002[b]
	Residual	784.272	22	35.649		
	Total	1402.080	24			

a. Predictors: (Constant), Rank of ArrowFla

b. Predictors: (Constant), Rank of ArrowFla, Rank of MedianEVS

c. Dependent Variable: Rank of Companies Average

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients			95.0% Confidence Interval for B		Correlations			Collinearity Statistics	
		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	19.126	2.922		6.545	.000	13.081	25.171					
	Rank of ArrowFla	-.402	.182	-.417	-2.202	.038	-.779	-.024	-.417	-.417	-.417	1.000	1.000
2	(Constant)	25.467	3.144		8.101	.000	18.947	31.987					
	Rank of ArrowFla	-.356	.154	-.370	-2.308	.031	-.676	-.036	-.417	-.442	-.368	.992	1.009
	Rank of MedianEVS	-.482	.149	-.518	-3.237	.004	-.790	-.173	-.552	-.568	-.516	.992	1.009

a. Dependent Variable: Rank of Companies Average

Regression: figures*2-back, median EVS

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	Rank of Nback[a]	.	Enter
2	Rank of MedianEVS[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: Rank of Figures

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.530[a]	.281	.252	7.091087	.281	9.778	1	25	.004
2	.652[b]	.426	.378	6.469528	.144	6.035	1	24	.022

a. Predictors: (Constant), Rank of Nback

b. Predictors: (Constant), Rank of Nback, Rank of MedianEVS

ANOVA[c]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	491.653	1	491.653	9.778	.004[a]
	Residual	1257.088	25	50.284		
	Total	1748.741	26			
2	Regression	744.226	2	372.113	8.891	.001[b]
	Residual	1004.515	24	41.855		
	Total	1748.741	26			

a. Predictors: (Constant), Rank of Nback

b. Predictors: (Constant), Rank of Nback, Rank of MedianEVS

c. Dependent Variable: Rank of Figures

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	6.795	2.817		2.412	.024	.993	12.596					
	Rank of Nback	.550	.176	.530	3.127	.004	.188	.913	.530	.530	.530	1.000	1.000
2	(Constant)	13.902	3.870		3.592	.001	5.915	21.890					
	Rank of Nback	.448	.166	.432	2.701	.012	.106	.790	.530	.483	.418	.937	1.067
	Rank of MedianEVS	-.391	.159	-.393	-2.457	.022	-.720	-.063	-.501	-.448	-.380	.937	1.067

a. Dependent Variable: Rank of Figures

Regression: median EVS*number-letter, experience in days

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	Rank of numletdiff[a]	.	Enter
2	Rank of ExDays[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: Rank of MedianEVS

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.421[a]	.177	.145	7.608109	.177	5.394	1	25	.029
2	.597[b]	.356	.303	6.869780	.179	6.663	1	24	.016

a. Predictors: (Constant), Rank of numletdiff

b. Predictors: (Constant), Rank of numletdiff, Rank of ExDays

ANOVA[c]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	312.250	1	312.250	5.394	.029[a]
	Residual	1447.083	25	57.883		
	Total	1759.333	26			
2	Regression	626.680	2	313.340	6.639	.005[b]
	Residual	1132.653	24	47.194		
	Total	1759.333	26			

a. Predictors: (Constant), Rank of numletdiff

b. Predictors: (Constant), Rank of numletdiff, Rank of ExDays

c. Dependent Variable: Rank of MedianEVS

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound	Correlations Zero-order	Correlations Partial	Correlations Part	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	8.387	3.012		2.785	.010	2.185	14.590					
	Rank of numletdiff	.437	.188	.421	2.323	.029	.049	.824	.421	.421	.421	1.000	1.000
2	(Constant)	14.686	3.654		4.019	.001	7.145	22.227					
	Rank of numletdiff	.425	.170	.410	2.501	.020	.074	.775	.421	.455	.410	.999	1.001
	Rank of ExDays	-.423	.164	-.423	-2.581	.016	-.761	-.085	-.434	-.466	-.423	.999	1.001

a. Dependent Variable: Rank of MedianEVS

Regression: experience in days*median EVS, arrow flanker

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	Rank of ArrowFla[a]	.	Enter
2	Rank of MedianEVS[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: Rank of ExDays

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.599[a]	.359	.333	6.716471	.359	14.000	1	25	.001
2	.710[b]	.504	.462	6.031393	.145	7.002	1	24	.014

a. Predictors: (Constant), Rank of ArrowFla

b. Predictors: (Constant), Rank of ArrowFla, Rank of MedianEVS

ANOVA[c]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	631.559	1	631.559	14.000	.001[a]
	Residual	1127.774	25	45.111		
	Total	1759.333	26			
2	Regression	886.269	2	443.134	12.181	.000[b]
	Residual	873.065	24	36.378		
	Total	1759.333	26			

a. Predictors: (Constant), Rank of ArrowFla

b. Predictors: (Constant), Rank of ArrowFla, Rank of MedianEVS

c. Dependent Variable: Rank of ExDays

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95.0% Confidence Interval for B Lower Bound	95.0% Confidence Interval for B Upper Bound	Correlations Zero-order	Correlations Partial	Correlations Part	Collinearity Statistics Tolerance	Collinearity Statistics VIF
1	(Constant)	23.193	2.659		8.724	.000	17.717	28.669					
	Rank of ArrowFla	-.621	.166	-.599	-3.742	.001	-.963	-.279	-.599	-.599	-.599	1.000	1.000
2	(Constant)	28.223	3.052		9.248	.000	21.925	34.522					
	Rank of ArrowFla	-.584	.150	-.564	-3.905	.001	-.893	-.276	-.599	-.623	-.562	.992	1.009
	Rank of MedianEVS	-.382	.144	-.382	-2.646	.014	-.680	-.084	-.434	-.475	-.380	.992	1.009

a. Dependent Variable: Rank of ExDays

Regression: companies average*arrow flanker, age, interaction arrow flanker and age

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	cRarrage, Rank of ArrowFla, Rank of Age[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: Rank of companies average

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	df2	Sig. F Change
1	.651[a]	.423	.341	6.205810	.423	5.135	3	21	.008

a. Predictors: (Constant), flanker*age, Rank of ArrowFla, Rank of Age

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	593.326	3	197.775	5.135	.008[a]
	Residual	808.754	21	38.512		
	Total	1402.080	24			

a. Predictors: (Constant), flanker*age, Rank of ArrowFla, Rank of Age

b. Dependent Variable: Rank of companies average

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics	
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF
1	(Constant)	18.077	4.236		4.267	.000					
	Rank of ArrowFla	-.393	.171	-.408	-2.305	.032	-.417	-.449	-.382	.874	1.144
	Rank of Age	.137	.167	.148	.823	.420	.358	.177	.136	.854	1.170
	flanker*age	.052	.020	.452	2.654	.015	.442	.501	.440	.947	1.056

a. Dependent Variable: Rank of companies average