Is working memory working in consecutive interpreting?

Ya-shyuan Jin, M.Sc.

A thesis submitted in fulfilment of requirements for the degree of Doctor of Philosophy

to

The University of Edinburgh

May 2010

Declaration

I hereby declare that this thesis is of my own composition, and that it contains no material previously submitted for the award of any other degree. The work reported in this thesis has been executed by myself, except where due acknowledgement is made in the text.

Ya-shyuan Jin

Abstract

It is generally agreed that language interpreting is cognitively demanding; however, to date there is little evidence to indicate how working memory is involved in the task, perhaps due to methodological limitations. Based on a full consideration of key components of interpreting, two series of experiments were conducted to explore how working memory might play a role in discourse and sentence interpreting. If working memory is implicated both in grammatical encoding into the target language, and in temporary storage of the discourse content, then higher demand in one function might compromise the other. Thus discourses that differ in word orders between languages could increase the processing load and leave less resource for memory maintenance, affecting recall performance. In Experiment 1, Chinese-English bilingual participants' memory performance was compared when they translated passages from Chinese to English and from English to Chinese, where the expected word order was either congruent or incongruent between source and target. Recall was not sensitive to word order or direction of translation. Perhaps surprisingly, memory for incongruent discourses was numerically better than that for congruent sentences. Experiment 2 showed that interpreting trainees performed just like the participants in Experiment 1 did, suggesting that memory performance was not modulated by translation direction in proficient translators. Experiment 3 explored the relationship between surface form transformation and recall. As discourse paraphrasing did not result in better recall than verbatim recall, it was concluded that the better memory performance for incongruent discourse interpreting suggested by Experiment 1 was not the result of active manipulation of word form or word order in interpreting. Finally, a free recall task among native English speakers showed that the incongruent discourses tested in earlier experiments were intrinsically more memorable than congruent discourses (Experiment 4). Despite this confound, this series of experiments highlighted the importance of comprehension in interpreting, but it did not rule out the role of working memory in the task.

The role of working memory in interpreting was further explored using on-line measures in Experiments 5-8. Experiment 5 replicated a self-paced reading study by Ruiz, Paredes, Macizo, and Bajo (2008), comparing participants' times to read sentences for translation to those to read them normally. The data showed that participants accessed lexical and syntactic properties of a target language in the reading-for-translation condition when resources were available to them. In order to explore the role of working memory in sentence interpreting, a dual-task paradigm was used in Experiment 6. When participants' working memory was occupied by a secondary task (digit preload), reading times were only different numerically between congruent and incongruent sentences. Crucially, reading times decreased as digit preload increased. Since there were no differences in the interpretations produced or in digit recall, it appears that participants were flexible in their resource allocation, suggesting that processing can be tuned up to optimise performance for concurrent tasks. Experiment 7 refined the procedure in the order of responses for the dual tasks but replicated the results of Experiment 6. A closer examination of participants' interpretation responses showed that devices that could reduce processing load in target language production may have been strategically employed. Finally, another set of sentences were used in Experiment 8 in an attempt to replicate Experiment 5. A failure to replicate the earlier findings suggested that working memory demand might differ for different syntactic structures in sentence interpreting.

All in all, this thesis shows that research in language interpreting benefits by taking a full account of the key components of interpreting. The use of on-line measures allowed us to take a ne-grained approach to the investigation of interpretation processes. It is proposed in this thesis that interpreting research may gain more insight from the data by incorporating some of the theories and methods typically used in research into language production.

Acknowledgements

The reaction of almost anyone who asked me how I changed careers was clear from their faces – wide-eyed and open mouthed. I often found it hard to cut my story short except by saying I just can't think of a topic more interesting than translation. To identify a new interest is one thing, but to invest all one's time, effort, and resources into it is quite another. It was risky. Somehow, my parents believed that I was a safe bet. I thank them for their consistent support throughout my study in the U.K. It would have been unthinkable, however, for me to embark on a PhD if Valerie Pellatt had not given me the encouragement to go to Edinburgh. I thank her for her guidance and fresh perspectives on my work. Although my parents were brave enough to send me abroad without knowing what to expect, the bravest of all must be my supervisors Robert Logie and Martin Corley. I am eternally grateful to them for believing in my potential and agreeing to work with me in the first place. Above all, their support was consistent and their insights were pivotal every step of the way in my research. I am most grateful to the Ministry of Education of Taiwan and Chiang Ching-Kuo Foundation who supported my research. My stay in Edinburgh would have been dull had I not met great friends in the department. Among them, Zhenguang Cai has been a good friend, a critical thinker and a basketball fanatic. I thank him for his advice and his generosity that kept me amused during our fussball sessions. I am also grateful to my brother and Shu-Chin both of whom kept motivating and inspiring me. I also thank Sneh Jaswal for her eagle-eyed proofreading and her help when I was almost overcome by despair. The progress in my thesis correction would have been a lot more challenging had I not been provided with great support by Billy Lee. The delivery of Chinese takeaway on a winter night by Billy and Thomas Wright is my most tasty memory. This thesis is also devoted to my late grand parents – grandpa was a role model of self-discipline, while grandma was always ready to crack jokes, and her practical intelligence has no rival in my family. Lastly, I'd like to acknowledge great companionship from the animal kingdom. Easty and Westy, my guinea pigs, did a great job in taking my stress away. Isabelle, Thomas' guidedog puppy, never failed to cheer me up, as a baby or as the big athletic pup that she has become.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Translation can be viewed as a series of processes including decoding information from a source discourse, retrieving information from the episodic memory where the message is temporarily stored, and formulating a sentence in a target language on the basis of the meaning of the message. However translations differ considerably in their delivery modality, availability of source discourse (transient in oral simultaneous interpreting (SI), permanent in written translation), time pressure, and so on. A theory based on one type of translation need not apply to the others.

In this thesis, the focus is on the type of translation that requires immediate delivery after a source discourse is provided. This choice of the interpreting mode that will be examined in this thesis was motivated by the intellectual curiosity accumulated ever since simultaneous interpreting was introduced in the end of Second World War and also by some interesting findings that could inform the interface between comprehension and production systems of bilinguals. The curiosity of how SI is possible stems from the complicated operations of overlapping comprehension and production demonstrated by interpreters in the media and conferences. This highly complicated task has led researchers to assume that interpreters might have extraordinary memory in addition to essential linguistic skills, such as large vocabulary and perfect command of the languages (Bajo, Padilla, & Padilla, 2000; Köpke & Nespoulous, 2006; M. Liu, Schallert, & Carroll, 2004). The topic of individual differences in working memory between interpreters and non-interpreters, however, gradually phased out because of some studies that discovered that professional interpreters were outperformed by interpreting students in some working memory tasks (Köpke & Nespoulous, 2006; M. Liu et al., 2004). This line of inquiry lent support to the argument that working memory is task-specific and that expertise seems to have a larger role than working memory in a higher cognitive activity like interpreting. Recently, another line of interpreting research appeared to coincide with that in the area of bilingualism. By considering language interpreting as a form of bilingual language processing, researchers have been able to address questions with more established theories and more tractable experiments (e.g., Macizo & Bajo, 2006; Ruiz et al., 2008). Following the findings in Macizo and Bajo (2006) and Ruiz et al. (2008), this thesis set out to further explore two main questions. First, does translation take place in a strict temporal order of source-language (SL) comprehension and target-language (TL) production or can processes overlap in time, possibly with rapid swapping between the two. Second, is working memory implicated in the process of interpreting?

Chapter 2

Literature Review

2.1 Interpreting – A complex operation calls for a full consideration of linguistic processes

When language interpreting is the subject of scientific inquiry, an operational definition of interpreting alone is insufficient if an interdisciplinary approach is taken. In addition to operationalising the definition of language interpreting, this section sets a theme that will run throughout the thesis by considering language comprehension, memory, language production, and cognitive resource, as four key components of interpreting. As an example, take *language production component* to appreciate why a full consideration of these four components is important. In the past, it is rarely, if ever, asked as to what is different and what is similar between interpreting and spontaneous language production. In other words, what characteristics/features does translation as a task have that spontaneous speech or other types of language production (picture description, dialogue, etc.) do not? The difference between them is apparent when translation and spontaneous speech are placed in the models of language production – apparently, the most salient and robust difference is the availability of material for the first step of language production – conceptualisation in Levelt's (1989) language production model. In spontaneous speech, a speaker generates a preverbal message on the basis of the goals of communication in a process called conceptualisation. In language translation, the material for translation comes from the message that is extracted by an interpreter from a source language(SL) discourse. Then he/she would generate a speech plan on the basis of the extracted message before executing this speech plan in a target language(TL). Therefore an interpreter relies to a great extent on his/her comprehension system in a translation task than a speaker would in spontaneous speech production. At times however, the source of material for interpreting includes not only the extracted message but also the original surface form, especially when the to-be-translated material is permanently present or when the representation of the surface form of an SL discourse remains highly activated in interpreters' working memory. This kind of scenario inevitably complicates the already complex process in interpreting and leads to a topic of debate as to whether SL comprehension and TL production can be construed as separate processes in language interpreting or if it involves an interaction between two languages during the stage of comprehension. This brief example using the *language production component* shows only one aspect of the complexity of language interpreting. In the light of potentially interwoven processes involved in interpreting, it is proposed that a full consideration of the four components of interpreting is critical in addressing the research questions in this thesis. After a brief section that operationally defines language interpreting, the overarching question of this thesis will be delineated.

2.1.1 Definition

Before giving the operational definition of translation for this thesis, it may be helpful to clarify the typological issues in translation studies and consider the purpose of translation. Translation is a generic term for language activities within and across languages and it can involve different modalities, e.g., sign language interpreting (Alexieva, 2002; Jakobson, 1966). With a narrower scope, translation refers to the reformulation of a written source text(ST) in a written target-language (Gile, 2004) whereas interpreting refers to oral translation by convention. De Groot (1997) proposed that translation should not be used to refer to both translation and interpreting because that could potentially obscure the fundamental differences in the cognitive processing of the two tasks. The typology of translation is relevant, but not critical to this thesis, because its topic concerns how translators process the ST before speech production and the thesis will also examine a potential interference of ST structure on target text formulation. By adopting Kade and the Leipzig School's view of interpretation as a form of translation (Riccardi, 2002, p. 77) and accepting the claim that there cannot be a general translation theory without the inclusion of interpretation (Riccardi, 2002, p. 80), the definition of translation for this thesis was kept intentionally wide, but not vague, to include all translation events whose source texts are either written or spoken. Translation and interpreting will be used interchangeably in this thesis, depending on the context of discussions, and the kind of task that is compared. The term 'translation' is used when comparing verbal communications that do and do not involve the switch of languages. Otherwise, 'interpreting' is used throughout the thesis. Although this thesis was inspired by research on SI, available theories and paradigms are not fit in studying SI in which two voice streams overlap in time. Contrary to SI, in consecutive interpreting (CI), translation only begins when a speaker gives his/her floor to the interpreter. The density of to-be-translated material depends entirely on the speaker. It can range from a sentence to a passage within a certain period of time. Under normal circumstances, interpreters are allowed to take notes in CI to aid their memory during interpreting. In order to study the role of memory in language interpreting by designing more tractible experiments, the interpreting mode of all tasks in the experiments of this thesis was consecutive. The only deviation in the experiments of this thesis from typical CI was that note-taking was prohibited.

In addition to the typology of translation, an ideal definition for translation should also consider translation as a communication act. It has been argued that translation is a goal-oriented activity (Broeck, 1991), the goal being 'to fulfill the same purpose in a different language as the original did in the language in which it was expressed' (Booth, 1958; Gutt, 1991). A definition of translation that only focuses on linguistic correspondence between languages without attending to the ultimate goal of translation would probably face difficulty in accommodating the translation of 'chalk' in English as 'writing soap' in Northeast Siberian Chukchees (Jakobson, 1966).

With typology issues and the purpose of translation considered, the operational definition of translation is a combination of Nida's (2003) definition of translation and Gutt's (1991) theory of relevance:

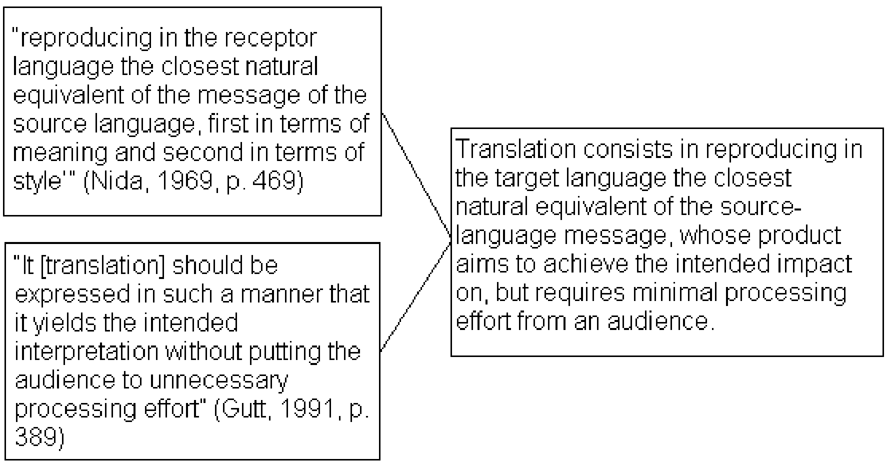


Figure 2.1: The operational definition of translation in this thesis

This definition not only preserves the essence of Nida's definition of translation, but also highlights the aspects of linguistic competence which contribute to optimal translation. The central feature of Gutt's theory of relevance is that the translator's job is not only reproducing 'the message' but also avoiding putting the target audience under unnecessary cognitive load. To avoid the latter, translation should not sound like a translation (Nida & Taber, 2003, p. 12)– words not only are put in correct word order according to the rules of a target language (TL), but also in an appropriate context. This kind of ability can be described in terms of translation competence, which will be introduced in the section, **Some facts about translators**, to characterise translators and translation.

2.1.2 Overarching question and approach

Interpreting typically involves two of the interpreters' languages: they either translate into their first language (L1) or into languages that were acquired later in their lives. One problem that is present in speech production of both L1 and second language (L2) is the "linearisation problem" (Levelt, 1989; Von Stutterheim & Nüse, 2003). The problem that a speaker faces is that decisions have to be made as to the order of expressions for a message that requires multiple speech acts. According to Levelt (1989), the solution to the linearisation problem depends on "the content of what is to be expressed on the one hand", and on the other hand, "there are general restrictions on working memory that induce a speaker to prefer one linearisation over another" (Levelt, 1989, p. 159). Conceivably, when a message is expressed in a less familiar language, e.g., the interpreters' second language, the problem may not only manifest in linearising ideas as in propositions within a discourse, but also in linearising the constituents within propositions (Pienemann, 2005). This is because when interpreters translate into their second languages, the factors that constrain their language production are not only their lexical knowledge (Section 2.3.3) but also language-specific procedures, e.g., grammatical encoding, that may have not been automatised. The demand of language interpreting on interpreters, however, can be compounded by another factor, memory.

As will be argued in the literature review, the definition of translation will lead on to a discussion of translation quality assessment with a focus on the essential requirement from a perspective of the translation users[1]. Accuracy and fidelity are two of the most important qualities for translation users. For an interpreter to meet such expectations, his/her memory for to-be-translated material plays a crucial role in a translation task, because when retrieved content is not complete, it is very unlikely that translation can be accurate and complete. Consider the sentences below:

1. 我們(we)上禮拜(last week) 在公園裡(in the park) 野餓(had a picnic)
We had a picnic in the park last week.

2. 警察(the policeman) 追趕(chased) 的(DE) 小偷(the thief) 跌倒了(fell)
The thief that the policeman chased fell.

The two examples show that the Chinese sentences differ from their English translation in their word orders. Example 1 exemplifies the difference of verb phrase complement word order between two languages. Example 2 demonstrated a robust word order differences of NP modification between the pre-positional modification in Chinese and post-positional modification in English. Because a sentence or a discourse is usually composed of more than one proposition, an accurate interpretation between two languages presupposes good memory recall for a to-be-interpreted text. But grammatical encoding at the stage of language production could be resource demanding. The first question the thesis attempts to address is whether discourse memory and grammatical encoding compete for working memory, resulting in a trade-off of memory performance when a to-be-interpreted material involves potentially more effortful grammatical encoding, e.g., in the examples above. In the light of the evidence suggesting that interpreters' two languages might interact during

[1] By translation users, they refer to the audience who relies on the translation service to understand discourses presented in a foreign language.

SL comprehension (Macizo & Bajo, 2006; Ruiz et al., 2008), the second question of this thesis concerns whether working memory plays a role in the process that implicates the interaction.

In order to address these two major questions, the task of interpreting needs to be examined in detail. A full consideration of the four components of interpreting in the beginning of the literature review will serve to highlight key topics of discussion in each of these components. The brief introduction of these four components of interpreting will be followed by a focused review of translation and interpreting models and theories. This review can reveal whether or not existing theories and models of interpreting or translation can be applied to address the major questions of this thesis. What will be argued is that these models/theories of interpreting were not developed specifically for addressing questions that were raised earlier. Therefore, by following the approach taken to study second language production by De Bot (1992, 2000) and D. Green (1993), it is suggested that a more viable approach would be adopting an established psycholinguistic model, e.g., language production model (Levelt, 1989), as a framework, which will serve as a guiding light in every step of the formulation of hypotheses for the experiments and in interpreting experimental results.

Since Levelt's (1989) model is central to the framework of this thesis, his production model will be briefly introduced before another key player in studying language interpreting, bilingualism, is discussed. They are key players because interpreting constantly involves two or more languages. So an understanding of how bilinguals have been characterised in their linguistic competence is fundamental if their data were to be used to constrain theories of interest. After this discussion, the four components of interpreting will be fully elaborated and the literature review section concludes by specifying the research questions and predictions in this thesis. Finally, tasks of discourse and sentence interpreting are used to test the hypotheses and the implications of the result for an understanding of working memory and interpreting processes are discussed.

2.2 Paradigms

2.2.1 Components of translation

As pointed out in the previous section, the conventional definition of translation has an obvious weakness in empirical study because translation is too complicated to be defined operationally, therefore it does not easily open up questions and formulate hypotheses that can be subjected to empirical testing. It is believed that in order to better understand translation, one needs to discover what it takes to achieve the goal stated in the definition of translation from a competence-performance point of view. Most critically, undertaking research into the subcomponents of translation is essential, since focusing only on one facet of a complicated task like translation can easily lose sight of theoretically more relevant aspects. Christoffels and De Groot (2005, p. 469) suggested that "by comparing novices and professionals on tasks that are supposed to tap into possibly relevant subskills, we can gain more insight into what cognitive subskills are in fact important for SI." Although it remains to be seen whether overall translation capability can be measured as the sum of the capability of individual subskills or whether it is indeed possible to decompose a task as complicated as interpreting (Setton, 2001), this approach is indispensable when there is not yet a clear psycholinguistic understanding of interpreting, and also when an interdisciplinary approach is adopted. Interpreting is defined and characterised by examining four components in the entire procedure of interpreting from input analysis to interpretation output. This approach is not entirely new. Gile (1991) proposed the Effort Model (EM), using equations to describe the resource required by different opeations involved in conference interpreting. Although this model can be used to describe both SI and CI, the discussion of EM here will focus on CI that was used throughout this thesis. Since the comprehension and reproduction phases are two discrete stages in CI, two equations were formulated for the comprehension phase and the reformulation phase respectively. In a typical CI task, the total resource required (TR) for comprehension equals the sum of the resource demanded by listening (L), note-taking (N), memory operation (M), and coordination (C).

Efforts involved in comprehension:

L + N + M + C = Comprehension

Total resource required for comprehension, TR_C :

$LR + NR + MR + CR = TR_C \geq T_A$

When it comes to the production phase, the TR is equal to the sum of the resource devoted to note-reading (Read), remembering (Rem), and production (P).

Efforts involved in production:

Read + Rem + P + C = Production

Total resource required for production, TR_P :

$Read - R + Rem - R + PR + CR = TR_P \geq T_A$

One major hypothesis made from the EM is the tightrope hypothesis. The combination of any two operations is thought to require more resource than would any single one, leading to a saturation of work load, thereby TR is at least equal to, or very often lager than, the total resource available. By examining the errors and omissions made in SI by a group of professional interpreters, Gile (1999) found that participants in their second round of test corrected errors and omissions made in the first round, but they also produced new errors/omissions, indicating that instrinsic difficulty of source text alone cannot account for the variability of interpreting performance (see (e.g., Matysiak, 2001) for similar findings). The EM has its limitation in its specificity and sophistication, admittedly, but the significace of this conceptual framework should be recognised in its potential in bridging interpreting research and scientific paradigms. Take the approach of this thesis to appreciate the EM, a deviation from typical CI is shown in the equation below and an elaboration will follow.

Efforts involved in comprehension: Listen/Read + M + C = Comprehension

Efforts involved in production: Rem + P + C = Production

Before elaborating on the equations, it is worth noting three components that (Gile, 2008a, p. 1-2) drew reference to working memory. The M component involves operations of "storage and retrieval of information related to the source and target speech in short term memory. They are collectively called 'Memory Effort', a concept distinct from but in many ways similar to the cognitive psychologists' working memory model(s)". The P component refers to producing a target speech, including self-correction and self-monitoring. The C component is the Coordination component which "manages attention allocation and shifts between the three[2]. Again it is distinct from but perhaps comparable to what Baddeley and Hitch have called the 'Central Executive' in their Working Memory model (Baddeley & Hitch, 1974a)". The removal of components note-taking(N) during comprehension and Read(note-reading) during reformulation from the original equation in the CI was deliberate to enable a more direct testing on the interplay between different components. When participants have to carry out CI without memory aid, it might be easier to observe performance difference that is attributable to the manipulations that was assumed to trigger the problems in operations of *Production* and/or *Remembering*. By putting the equations of *Comprehension* and *Reformulation* on a time line, four

[2]In the adapted equations for the purpose of this thesis, the coordination effort would shift between the other two components rather than three.

key components (comprehension, memory, production, and resource) were identified from the formula below.

Listen/Read + M + C $\rightarrow$ *Remembering* $+ P + C$

The justfication for singling out four, instead of five, components is that there is no clear reason as to why Listening or Reading component must be separated from Memory in the comprehension component if the notion that memory is intrinsic to comprehension is adopted (Kintsch & Van Dijk, 1978). This point is discussed later in the full elaboration of translation components. Each of the components will be introduced in turn in the following section.

So there is a clear influence of EM on the approach taken in this thesis and there exist similarities in hypothesised key components of interpreting. But a major step forward taken in this thesis might be the attempt of discussing the key components in the framework of speech production by Levelt (1989). By regarding the representation constructed during SL comprehension as the ingredient for speech planning and grammatical encoding in a TL, and by assuming working memory as a workspace (Baddeley & Hitch, 1974b) where bilingual lexicons interact and as a workspace which temporarily stores and process information, it could lead to productive discussion and testable hypotheses.

As indicated earlier, these components set a theme that runs throughout the whole thesis. The result of this section will be the precursor for the more detailed discussion in section 2.4 and it will lead to the research questions in section 2.4.5.

Memory Component

One critical but rarely mentioned topic in interpreting research is to what extent is production in translation a memory recall task? If J. K. Bock (1996, p. 400) was correct by likening immediate recall to "producing a response by assembling highly activated linguistic elements, using the mechanisms of production to do so," maybe translation only differs from recall in the language used for production. But what is most intriguing must be the question as to what the "highly activated linguistic elements" are in the context of translation, and how differently the memory content of a discourse is re-represented verbally in different languages. The two questions can develop into two parallel lines of inquiry. On the one hand, one has to distinguish the information left in memory after parsing a sentence for immediate recall and the information left in memory for immediate translation, given that subjects are told in advance of the purpose of the discourse comprehension. In other words, is information bound differently in comprehension for understanding and for translation? On the other hand, one can contrast the mechanism underlying conceptualisation in spontaneous language production and the mechanism subserving memory reconstruction before its content is vocalised by a production system. Noticeably, the component of interest in common between two lines of inquiry is memory.

But can this line of inquiry tell us anything except that languages are different between a recall and a translation task? The difference may seem irrelevant at first glance, however, the comparison will be informative when recall and spontaneous speech are placed in the framework of a language production model.

According to Levelt (1989), the first stage in spontaneous language production is conceptualisation. During conceptualisation, a speaker generates a preverbal message according to his/her intention of communication. The preverbal message will be converted into a speech plan by selecting lexical items and applying grammatical and phonological rules. And finally this speech plan is converted into speech by the Articulators. In spontaneous speech, preverbal messages are generated *first hand* by a speaker him/herself, whereas in sentence recall, material must be retrieved in the first place to construct a mental representation of sentence memory. But does it take an extra stage to generate preverbal messages at this point or does the mental model of the reconstructed memory have all the information needed for the next step of production, grammatical encoding? J. K. Bock (1996, p. 415) believed that this remained an unresolved issue but speculated "elicited recall of these messages resembles normal speech in so many ways that it has seemed reasonable to suggest that what is called 'reconstruction' in memory for linguistic materials is in many respects equivalent to the process of language production." To the extent that her comment holds, it may be reasonably safe to hypothesise that a separate procedure does not need to be assumed in a sentence recall task up to the stage where preverbal messages are generated.

Even though the same assumption may also be true before a translator starts speaking in a TL, speech performance in many aspects may not be compatible between translation and recall because of a thorny issue, the relationship between language and thought. It is uncertain how dependent memory recall is on language. If recall from memory is largely language-dependent (Marian & Neisser, 2000), it is conceivable that translation will be constantly influenced by two factors. One is language proficiency of the translator whose effect is manifest in real-time problems during translation, such as word retrieval. The other is the set of properties of source and target languages. This factor is particularly relevant if a dominant language is used for retrieving memory and in planning utterances along the course of translation. There could potentially be an interference due to the structural differences between two languages.

In summary, the *memory component* of translation taps into a few fundamental questions related to human cognition and also points to a unique demand which sets translation apart from spontaneous speech production and sentence recall. It also opens up theoretically significant questions that will be addressed in this thesis – the relationship between memory and translation performance.

Comprehension Component

The *comprehension component* of translation has been a topic of debate among translation theorists. The topic concerns the information that a to-be-translated sentence may provide a translator. Of great theoretical interest, the question can be rephrased as whether input analysis in translation leaves only meaning but completely erases all traces of linguistic information of source language discourse. According to more than twenty years' history of language production studies (Pickering & Ferreira, 2008), the latter does not appear to be the case. This viewpoint is elaborated later in section 2.4.3.

The importance of the *comprehension component* is self-evident in that translation opens a new window to understanding human languages, particularly cross-cultural communication, since there is a possibility that the conceptualiser of language production may not work language-independently as has been thought (De Bot, 2003). Critically, the component of comprehension to some extent decides the ease with which information is encoded, memorised, and activated during the encoding phase of translation. How this derived information becomes optimally utilised by an interpreter is therefore strongly related to the *memory component* which is inseparable with the *resource component*. As mentioned earlier that source of material for interpreters to generate perverbal messages for TL production comes not only from his/her comprehension of a given discourse, but also from the surface form that still remains highly activated. So under some circumstances, it has been thought that parallel translation (De Groot, 1997) could occur, whereby participants access the lexical and/or syntactic information of a TL during SL comprehension. In other words, processes in translation are not strictly serial as was thought by *vertical translation* theorists. This position was questioned by the other camp of translation theorists representing *sequential translation* who argued that translation is no different from monolingual language comprehension and production because all translation is meaning-based (Seleskovitch, 1976), and suggests that the surface form of the source discourse decays when its meaning is extracted. This theory implies that when the processes in language comprehension and production are understood, it would easily explain how translation is carried out because comprehension and production are two discrete and serial procedures in translation. In the light of their theoretical and pedagogical significance, it is therefore not surprising that researchers have been keen to test the hypotheses of parallel and sequential translation, which is one of the themes of this thesis.

Grammatical Encoding component

In addition to the memory and comprehension components, there is a third and the most controversial one, the *grammatical encoding* component. In Levelt's (1989) model, grammatical encoding takes place in a Formulator where a preverbal message is converted into a speech plan. The same procedure is also assumed by DeBot's bilingual production model (De Bot, 1992). But when these models are applied to translation, it has been asked whether grammatical encoding during translation occurs exclusively in the phase of language production in a target language. Recent evidence (Macizo & Bajo, 2006; Ruiz et al., 2008) suggested a possibility of 'parallel translation' where the production system of a TL may have been accessed by interpreters during comprehension. The significance of the grammatical encoding component of interpreting can be evaluated in several aspects. One aspect concerns the implication of the discovery of evidence for parallel translation, which could lead theorists to re-think the models of language production, particularly when most people in the world can speak more than one language. Another aspect that is a major theme in this thesis is how memory retrieval and language production interact. Most memory tasks require verbal responses, e.g., oral serial and free recall, but little attention is paid to the very *act* that verbalises retrieved memory. Similarly, in language production studies, the focus has been the process of grammatical encoding, but little is known about how a speaker acquires the material for expression. As argued in the section of the memory component, this is directly relevant to language interpreting, as interpreters might be involved in free recall in a target language. Specifically, it is both theoretically and practically significant to explore whether current theories of memory and language production could shed some light on the question of whether linguistic structural differences constrain memory retrieval.

Resource Component

Apart from components of comprehension, memory, and grammatical encoding, another component that has been assumed to have a crucial role in interpreting is the cognitive resource that coordinates processes involved in translation in order to achieve the goal under constraints of time pressure and resource limitation. Although it is ideal to conceive the resource in a wholistic perspective in which the resource includes translators' crystallised and fluid intelligence, the focus of resource use in highly complex activities such as language interpreting has been short-term memory or working memory. While acknowledging the indispensable role of prior knowledge stored in long-term memory in language interpreting, this thesis focuses on how working memory is implicated in the other three components of translation.

Taken together, the four components were thought of particular interest and significance to translation theory development because these are the processes that appear to interface comprehension and production systems of a multilingual brain in cross-language communication. They give translation a unique identity whose procedures bear similarities to those of within-language comprehension and production, and at the same time they also help identify and formalise a testable hypothesis that is centred around the linearity issues in language processing and resource allocation during translation. Each component raises one or more questions central to the formulation of research questions in this thesis. In order to arrive at testable research questions and account for the data acquired later on, a theoretical framework is required. To start with the review, existing theories and models of translation/interpreting are discussed at the beginning of the next section. After the review of theories/models, it is considered whether there is a translation/interpreting model that provides a suitable framework for addressing the core questions of the thesis.

2.2.2 Translation Theories and Models

A need for a systematic and empirical approach to studying translation and interpreting was recognised well before language interpreting established its professional status and a research subject in its own right. Sanz (1931) described the work and abilities of conference interpreters with a focus on their cognitive abilities, stress factors and training needs. The demand for translators and interpreters surged immediately after the Second World War, presumably triggered by the recruitment of interpreters for the Nuremberg Trial and later for the United Nations. There was not much research going on in the 1950s. But during the 60s and 70s, translation research not only grew in quantity, it also enjoyed a great diversity in approaches when Henri Barik, Frieda Goldman-Eisler, and David Gerver offered their insights from psychological/psycholinguistic perspectives. During the same period of time, Chomsky's (1965) generative transformational grammar inspired Nida's (1969) deep structure approach. In addition, linguistic analytic and functional approaches resulted in theories advanced by Stein (1980), Diller and Kornelius (1978) and Hönig and Kußmaul (1982). For the purpose of the following discussion, these theories and models will be roughly categorised as linguistically- or psychologically-motivated, and they will be evaluated with respect to how successful they are in explaining translators' performance by taking account of individual differences in bilingual competence.

The frequently cited *Science of Translation*(Nida, 1969) describes translation as an activity involving three sequential stages: analysis, transfer, and restructuring (Figure 2.2). The analysis of ST covers grammatical, referential-semantic, and

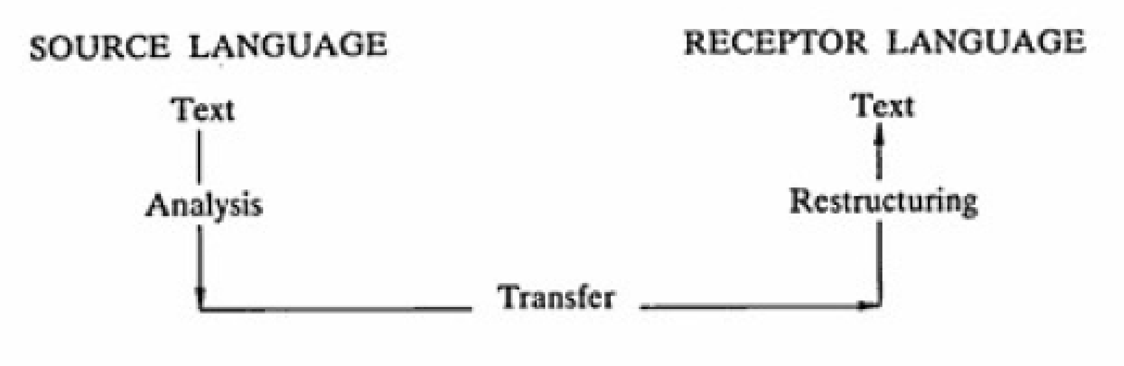


Figure 2.2: Nida's model of translation process, adapted from Lörscher (1989).

connotative aspects. It is the grammatical analysis that receives the strongest criticism. Surface structures of ST are transformed into kernels or deep structures in the grammatical analysis, so that complex structures are transformed into simpler ones. These kernels are then transformed into surface structures in a target language. Although Nida (1969) mentioned that translators must take account of receptors' interest, and the intended effect of a source text must be well calculated and realised in the target language surface structures, the notion of kernel was not clearly defined. The abstractness of these sub-processes and their strict temporal sequentiality attracted criticism from Hönig and Kußmaul (1982) who suggested that transfer can overlap with the phases of analysis and synthesis. They hint at the possibility of direct segment association between languages, i.e., it is only when comprehension is effortful that ST has to be analysed in chunks of more elementary units. For its lack of clarity in defining each stage of his model, Nida's (1969) model does not seem to meet his own definition of theory:

> A theory should be a coherent and integrated set of propositions used as principles for explaining a class of phenomena. But a fully satisfactory theory of translating should be more than a list of rules-of-thumb by which translators have generally succeeded in reproducing reasonably adequate renderings of source texts. A satisfactory theory should help in the recognition of elements which have not been recognized before, as in the case of black holes in astrophysics. A theory should also provide a measure of predictability about the degree of success to be expected from the use of certain principles, given the particular expectations of an

> audience, the nature of the content, the amount of information carried by the form of the discourse, and the circumstances of use (Nida, 1991, p. 20).

Kade (1968) defined translation as the process of recoding from one language to another and distinguished translation and interpreting in terms of the time span in which target-language discourse can be checked and corrected. Within the framework of a communication theory, Kade (1968) designated the translator a role of message receiver (E) who is involved in a communication using a source language (SL) with the message sender (S) in the first stage of translation. In the second stage, the translator who has a different identity (U) recodes language from L1 to L2. In the final stage the translator takes up another role as a message sender (S) and initiates the communication with the target-language receiver (E') by means of a target-language (TL) (Figure 2.3).

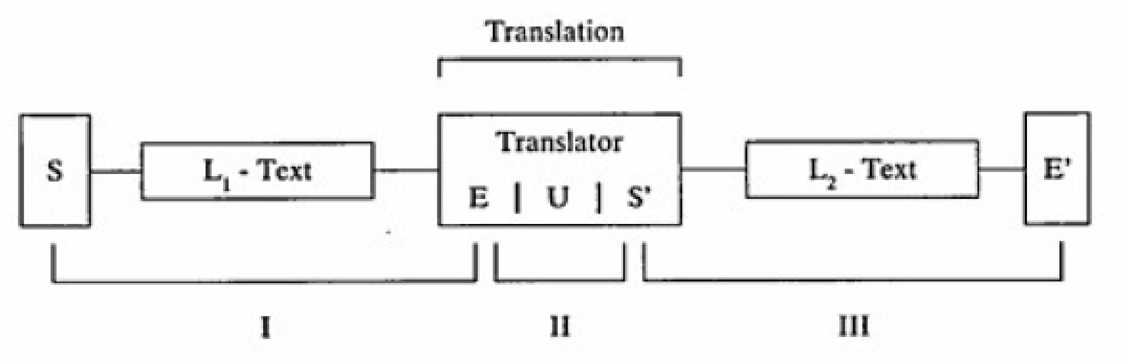


Figure 2.3: Kade's model of translation process, adapted from Lörscher (1989).

The generality of Kade's model was challenged by Lörscher (1989) for its unclear definition of how the recoding between languages took place, which rendered the interdisciplinary approach, including linguistics, psychology, and neurophysiology very difficult, if not impossible. Similar to the weakness of Nida's model, Kade's model cannot capture the mental process of translation, and it has the same issue of assuming the three stages as discrete and successive in operation.

Diller and Kornelius's (1978) model bears great similarity to Kade's but differs in the sophistication of the component of meaning. Diller and Kornelius (1978) proposed that meaning has seven components and an SL text was only equivalent to a TL text when they were equivalent semantically, pragmatically, and stylistically. Their theory is therefore a product-oriented one, but what was troubling Lörscher (1989) was that it was possible to have translation that contradicted their rules and that the assumption of equivalence in meaning between SL and TL was also problematic.

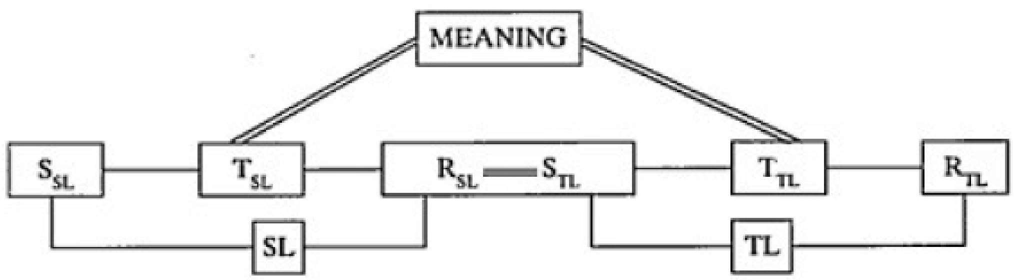


Figure 2.4: Diller's model of translation process, adapted from Lörscher (1989).

Stein's (1980) model (Figure 2.5) features communicative intention (I), function (F), situation of the receiver (Sit) and also the textual knowledge (Text). A translator has to deduce the intention of the SL sender first, and continues by weighing up the receiver's situation and textual knowledge, which should lead to selection of appropriate TL signs (TL S) that might produce a communicative effect (F2) similar to the effect felt by the translator him/herself (F1).

Because this model features selection and implies the aspect of decision-making in translation process, one weakness in Stein's (1980) model, pointed out by Lörscher (1989), is the mechanism for selecting appropriate TL signs when there are more than one sign with similar communicative function. Another weakness is directly relevant to the discussion in section 2.3.3. Lörscher (1989) suggested that Stein's model is an idealised and prescriptive model. "As an ideal translator does not exist, and as every translator has to work under non-ideal circumstances (i.e., a limited

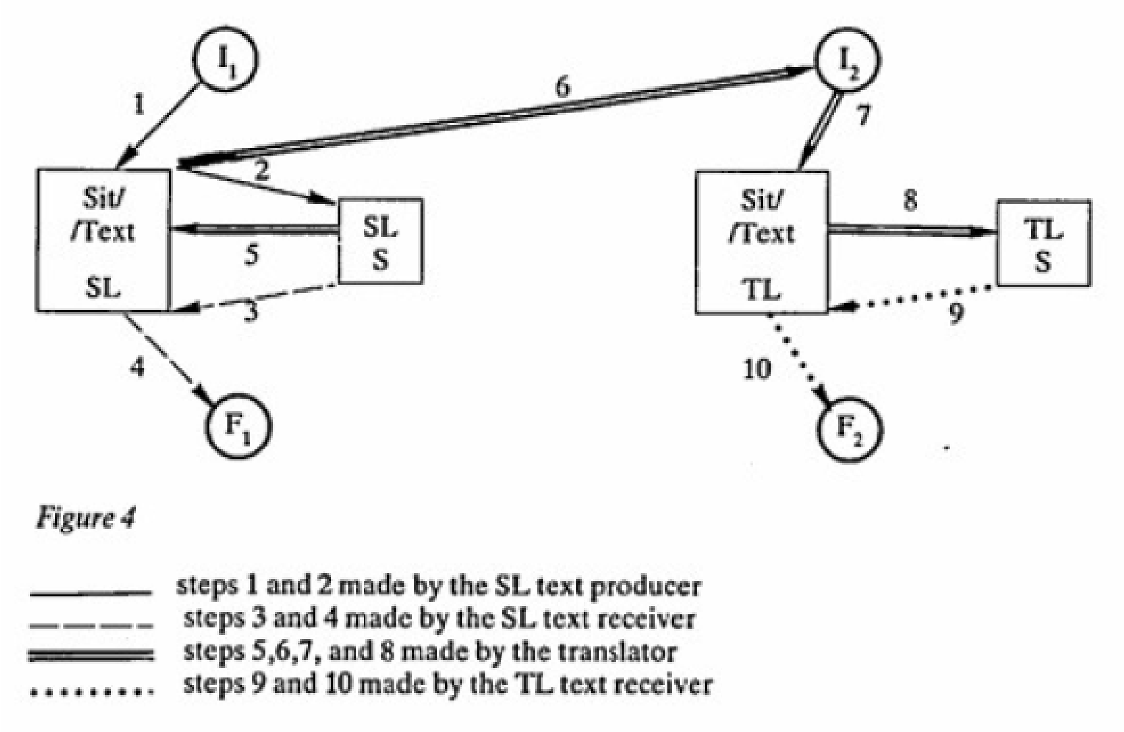


Figure 2.5: Stein's model of translation process, adapted from Lörscher (1989).

memory capacity, an incomplete Sit- and /or Text-knowledge, a limited availability of SL and/or TL signs, deficient mental mechanisms, translational experience, etc.), the model cannot adequately capture the real, unvarnished translation performance" (Lörscher, 1989). Apart from its idealised sketch of the model, another issue is the assumption of serial processing, shared by the models discussed above (Diller & Kornelius, 1978; Kade, 1968; Nida, 1969). It is noteworthy that Lörscher (1989) highlighted that translation performance is constrained by multiple variables, including translators' competence at the lexical level, among other things. Generally speaking, these linguistically motivated theories reflected the researchers' intuition, but then this is not unusual in the early phase of theory development, as Holmes (1987) pointed out that "translation theory, for instance, cannot do without the solid, specific data yielded by research in descriptive and applied translation studies, while on the other hand one cannot even begin to work in one of the other two fields without having at least an intuitive theoretical hypothesis as one's starting point." As for the question whether these models are capable of explaining translators' performance, Lörscher (1989) concluded his critical review by suggesting that none of the models he reviewed can account for the psychological reality of translating, specifically because they were difficult to subject to empirical test.

In psycholinguistically motivated theories, intuition also played a role to different degrees in researchers' development of theories focusing on the mental processes during translating and interpreting. A dominant theory in the 70s was the frequently cited *Théorie du Sens* (Theory of Sense) championed by Seleskovitch (1976). The essence of this theory is the notion of *deverbalisation* of SL material when comprehension is completed. The result of this procedure is a non-linguistic *sense*. Figure 2.6 shows two routes with which messages may be transferred between languages. Seleskovitch and Lederer (1992) emphasised that re-expression for an SL material in a TL must be based on the sense (*interpreting route*), but not on the surface form (*transcoding route*). She cited Ciceron and Hierouymus [3] who argued that word-for-word transcoding is not ideal since it usually renders translation unintelligible. But he also added that any translation entails transcoding to a certain degree, because this is inevitable in translating novel terms that interpreters have never encountered before.

The context in which Seleskovitch's (1976) model has been cited in recent interpreting studies (Christoffels & De Groot, 2005; Macizo & Bajo, 2004, 2006; Ruiz et al., 2008) was hypothesis testing between horizontal and vertical translation (De Groot, 1997), which are also known as lexical associated and conceptually-mediated translation (Paradis, 1994b) respectively. Seleskovitch's (1976) and Paradis's (1994b) models will be discussed side by side.

[3]Marcus Tullius Cicero (106-143 BC) was a Greek politician and laywer. Hieronymus (347-420 BC) was a Greek priest and physician.

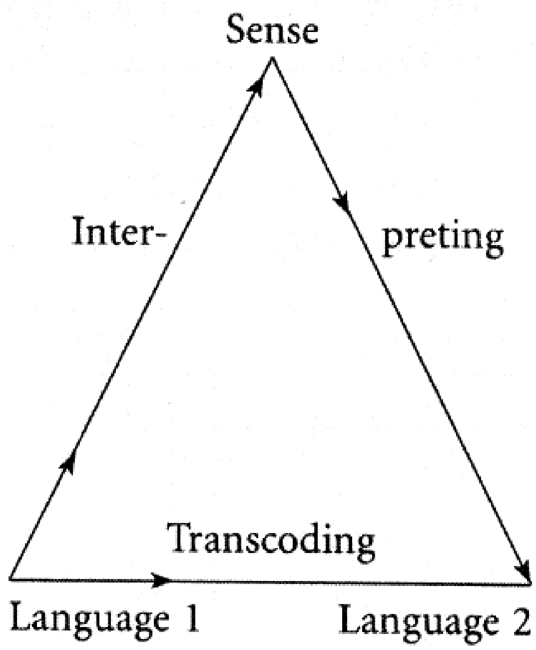


Figure 2.6: Seleskovitch's model of interpreting process, adapted from Seleskovitch and Lederer (1984, p. 185).

Before elaborating on Paradis's (1994b) model, one might suggest that Seleskovitch's and Paradis' models are more similar than they differ in their appearances. Both models show unidirectionality in the information flow, and they both allow transcoding. The only difference is that Paradis's (1994b) model incorporates and explicates the concept of bilingual lexical recognition and production. Seleskovitch's (1976) model however was questioned by many scientific-minded interpreting educators in the School of Modern Languages for Interpreters and Translators (SSLMIT) conference [4] in 1986, as the model suffers drawbacks similar to those linguistically-motivated models, in that these could not be subjected to empirical testing. Therefore the early 1990s saw the *empirical turn* (Pöchhacker, 2008) when Gile (1990) called for stringent empirical research on interpreting, along with people like Barik, Gerver, Morser-Mercer, Mackintosh, Pinter and Stenzl.

[4]SSLMIT was the international symposium on conference interpreter training organised by the school of Translating and Interpreting, University of Trieste. Note that SSLMIT is the abbreviation for the *Scuola Superiore di Lingue Moderne per Interpreti e Traduttori.*

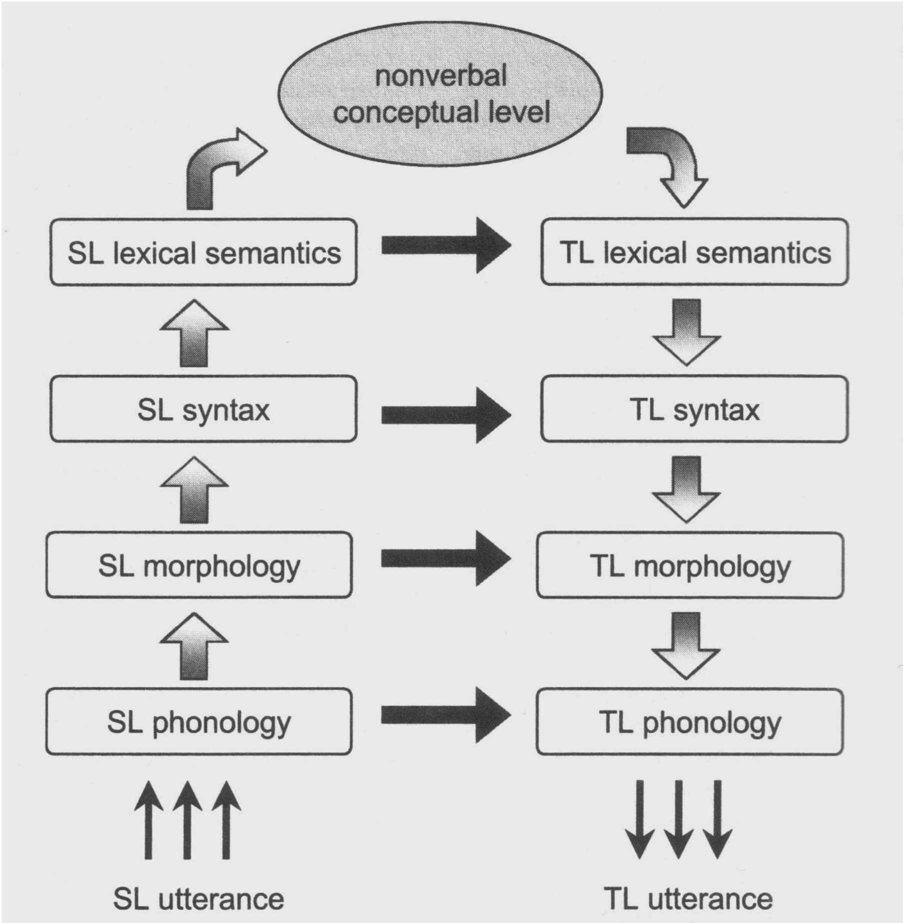


Figure 2.7: A model of simultaneous interpreting based on Paradis (1994). The light arrows represent the route of conceptually-mediated translation whereas the dark arrows depict the transcoding strategy, adapted from Christoffels and De Groot (2005, p. 460).

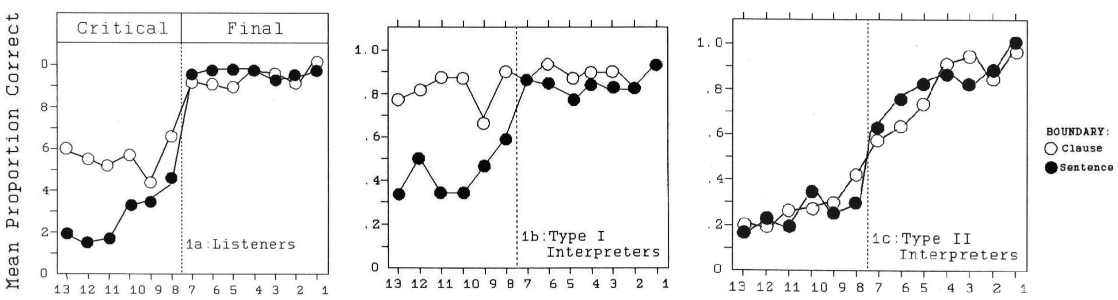
Nevertheless, Seleskovitch's (1976) model might be best recognised for its didactic value (Gile, 2008b) and Paradis's (1994b) model has both translation strategies. Vertical translation (also known as sequential or conceptually-mediated translation) involves full comprehension of the source discourse in a way similar to within-language comprehension. TL production is then based on the extracted nonverbal representation. According to Paradis (1994b), horizontal (also known as parallel or transcoding) translation involves direct transposition of SL and TL units, and this transcoding can take place at any of these levels shown in Figure 2.7.

Is there any evidence that can support Paradis's (1994b) model? Two approaches are available in testing vertical vs. horizontal translation strategies. One is to test the predictions made from the notion of deverbalisation: whether or not processing SL discourse leaves any trace of SL activation after the comprehension process is completed. If *deverbalisation* is the hallmark of vertical translation, translators should preserve no memory of the surface structure of SL discourse after SL comprehension. The other approach is detecting any sign of TL interference on SL processing: whether accessing SL material leads to the activation of TL properties. If vertical translation holds, TL should only become active when translation output is required, i.e., performance of SL processing should be the same irrespective of the purpose of SL comprehension, be it comprehension or translation. If horizontal translation is implicated, i.e., TL properties become activated side-by-side with SL processing, the model predicts an interference effect on the SL processing, e.g., reading time. Understandably this approach will have to address the question as to why TL activation interferes with, but does not facilitate, the SL processing. Nevertheless, either way, the performance of SL processing for translation should differ from that of the baseline task, SL processing for comprehension.

As regards the memory trace approach, Isham (1994) followed Jarvella's (1971) paradigm to examine whether or not instruction of reading and the location of boundary, at which point participants were interrupted and prompted for verbatim sentence recall (within sentence vs. across sentence), influence the proportion of correctly recalled words. Jarvella (1971) reasoned that limited memory capacity constrains the size of unit of processing, therefore listeners have to segment a speech into natural units, e.g., sentences or clauses, and arrive at an integrated semantic representation before their memory is overloaded. The key assumption was that syntactic segmentation may be immediately followed by its semantic integration, and once semantic integration is completed, some loss of perceived syntactic structure could occur. This assumption would lead to the prediction that the level of integration correlates negatively with the amount of verbatim recall of the surface form of previously heard speech segment, i.e., "speech which has been only segmented may be recalled better than partially interpreted speech, and both might be expected to be remembered more accurately than speech already fully interpreted, thus the accuracy of immediate verbatim recall may potentially distinguish among several degrees of processing" (Jarvella, 1971). His results suggested that listeners' verbatim recall was a function of the location of syntactic boundaries, and there was a recency effect. For instance, in the example below, the clause the union had even brought in outsiders was recalled better than its previous clause, suggesting that surface structure is most readily retrievable for only the most recent clause. Those clauses that were further away from the point where participants were prompted for recall were rarely recalled word-for-word but largely paraphrased by the participants. He also showed that the second last clause to stack the meeting for McDonald was recalled better in sentence 3 than 4, indicating different degrees of semantic integration: sentences appeared to be less fully integrated at a within-sentence clause boundary than at a between-sentence clause boundary, hence better preservation of surface structure for verbatim recall.

3. The confidence of Kofach was not unfounded.
To stack the meeting for McDonald, the union had even brought in outsiders.
4. Kofach had been persuaded by the officers
to stack the meeting for McDonald. The union had even brought in outsiders.

Isham (1994) extended the same idea to language interpreting, particularly to test Seleskovitch's (1976) *deverbalisation.* Interpreters and monolingual controls read passages as if they were going to translate and comprehend them respectively. During reading passages, they were interrupted irregularly and were required to write down as much as they could remember from the passage. If interpreters processed SL passages in the same way as controls did, it was expected that they would not differ in their recall performance. And if Seleskovitch's (1976) *deverbalisation* was implicated and no trace of surface structure of SL material remained, there would be no group difference or clause boundary effect. Overall, interpreters performed poorer than controls, but interpreters' performance patterns were grouped by Isham (1994) into type I and type II[5]. Figure 2.8 shows that whereas controls and type I interpreters were sensitive to the location of clause boundary, type II interpreters were not affected. The patterns of data for controls and type I interpreters replicated Jarvella's (1971) results: verbatim recall for the final clause was superior to recall for words in a previous clause. Consistent with Jarvella (1971), the effect of clause boundary location among type I interpreters, i.e., words in clause To stack the meeting for McDonald of version A were better recalled than when the same words fell in a different sentence, as in version B. This indicated that the surface structure prior to a within-sentence boundary had to be held temporarily to be integrated with the rest of the sentence content later, therefore it was more accessible. In other words, type I interpreters' word recall appeared to show traces of SL surface structure, which was contrary to Seleskovitch's (1976) full deverbalisation account.



(a) Listeners (b) Type I Interpreters (c) Type II Interpreters

Figure 2.8: The proportion of words correctly recalled as a function of location of clause boundary. The Y-axis of each line graph represents the position of each word in a to-be-recalled sentence, the larger the number, the further away it is from the sentential final position. Adapted from Isham (1994).

[5]These types were not labelled but served to distinguish two performance patterns.

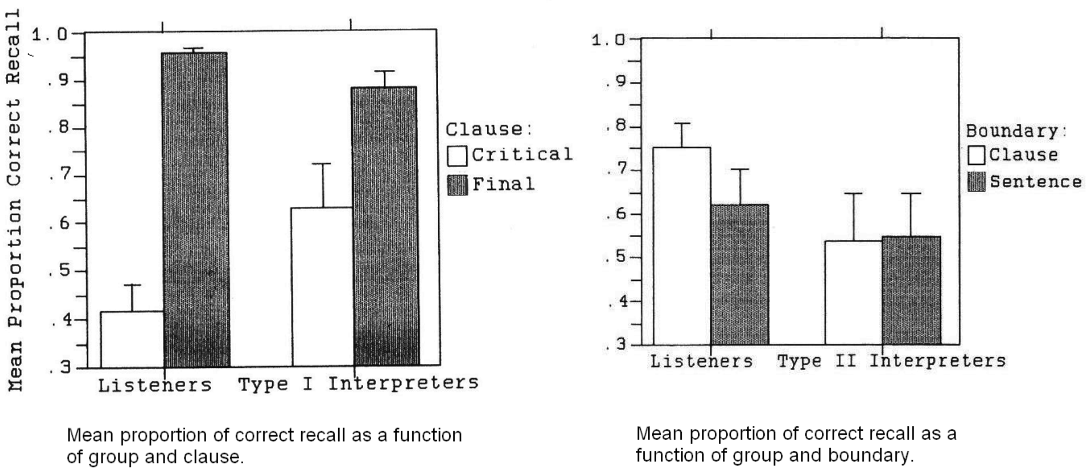


Figure 2.9: The probability of word correctly recalled as a function of location of clause boundary, adapted from Isham (1994).

When examining Figure 2.8 together with the right panel of Figure 2.9, type II interpreters clearly performed rather differently from the other two groups. The most salient features are the absence of a clause boundary effect and a generally poorer recall compared to that of controls (listeners). Isham (1994) suggested that these two characteristics seemed to indicate that "type II interpreters processed incoming sentences in a manner that drew attention away from their surface form, thus leaving little trace in memory of the original parsing of words into clause or sentence units." Therefore, type II interpreters' processing was more consistent with Seleskovitch's (1976) deverbalisation account. The only concern in this paradigm is that the task did not really involve overt translation in the TL, French (since verbatim recall in English was required), leaving Isham's (1994) discussion on participants' strategy deployment not all that convincing. The requirement of verbatim recall, emphasised in the task instruction, could have undermined the internal validity of this paradigm. As Isham (1994) admitted "It may also be that the type I interpreters normally use a meaning-based approach, but switched strategies for the purposes of this experiment, in an effort to be better able to meet the demands of the verbatim recall task."

As regards the other approach to testing the sequential vs. parallel translation hypotheses, the objective would be to detect whether there is an impact on SL processing that can be attributed to the activation of co-activated TL properties. Recently Macizo and Bajo (2006) directly tested these hypotheses and showed that translators as well as fluent bilinguals accessed TL lexical properties in a self-paced reading study. In two of their four experiments, participants were presented with Spanish sentences word-by-word, after which they were required to repeat sentences aloud in Spanish or translate them into English. Among their professional translators, there was a significant cognate effect[6], i.e., participants' reading time (RT) was shorter when the Spanish words were English cognates than when the words were not cognates. Crucially, the cognate effect was present only in reading-for-translation but not reading-for-repetition conditions. The same results were replicated in a group of fluent Spanish-English bilinguals in the same article. These results were interpreted as evidence for parallel lexical activation in a translation context. And it was also supported by the cognate effect found in an ERP study with similar design (Ibáñez, Van Hell, Macizo, Witteman, & Bajo, 2008). More recently, Ruiz et al. (2008) tested these translation hypotheses further by capitalising on a well-studied effect of lexical frequency in reading – low-frequency words tend to be fixated on or read longer (Rayner & Duffy, 1986). In their ingenious design, Spanish words that are matched for frequency were divided into high and low-frequency types according to the frequency of their English translation equivalents. The task procedure was otherwise similar to that of Macizo and Bajo's (2006). Ruiz et al. (2008) replicated the effect of reading purpose, but most interestingly, there was a frequency effect – the reading time of Spanish words were longer when their English translation equivalents were low-frequency words than when they were high-frequency words. Since the frequency effect was only found in reading-for-translation condition, this interaction between reading purpose and

[6] Cognates are pairs of words with the same meaning in two languages and similar pronunciation and/or spelling (Ardila, 2003).

lexical frequency provided further support for the parallel translation hypothesis. This finding is also consistent with the co-activation of semantically related lexical items in a non-target language when bilinguals were presented with target language items in a lexical decision task or presented with pictures for picture naming in bilingual studies reviewed in section 2.3.3. Macizo and Bajo (2006) suggested that the longer RT may have resulted from an extra cost the parallel translation incurred during TL lexcial retrieval on top of the resource needed for general comprehension.

The same inquiry has been extended from the lexical level to the syntactic level. In Ruiz et al.'s (2008) second self-paced reading experiment, the flexible word order of Spanish adjectival phrases was used to test the translation hypotheses. Compared to the rigid Adj-Noun (A-N) order of English adjectival phrases, Spanish adjectives can be placed on either side of the nouns they modify, although there is a subtle pragmatic difference between two usages. Each of Ruiz et al.'s Spanish sentences had two versions which differed only in the word order of the adjectival phrase. The congruent version had the same word order as its English translation (e.g., verde césped - green lawn), whereas the incongruent version reversed the adjective and the noun (e.g., césped verde - lawn green), hence a mismatch of word order. Their rationale was that if translation involves parallel TL syntactic activation, reading times would be longer in incongruent than congruent conditions, and this reading time difference would be present in translation condition only. On the contrary, if reading for translation did not implicate parallel TL syntactic activation, there would be no difference in RT. Their results confirmed the effect of reading purpose where the reading time in reading-for-translation conditions was reliably longer than that in reading-for-repetition, suggesting that resources had been allocated in accessing TL lexical information. Most importantly, their participants were slower in reading when word orders were incongruent than when they were congruent, but only in the reading-for-translation, not the reading-for-repetition condition. This interaction led Ruiz et al. (2008) to conclude that the congruency effect may be driven largely by the activation of TL syntax and perhaps also by searching for syntactic matches in the TL, both of which demand working memory. The effect of word order congruency between SL and TL has also been reported in a study using eyetracking (Sjørup, Jensen, & Balling, 2009) in which professional translators read Dutch sentences for later translation into English while their eye movements were recorded. Compared to the fixation duration on Dutch phrases with subject-verb word order, longer fixation duration was observed before their participants translated phrases with verb-subject word order, which is incongruent with the subject-verb order of its English translation.

In sum, current evidence from studies of bilingual lexical access and self-paced reading support the transcoding strategy in language translation, but the evidence cannot be taken to rule out the vertical translation strategy. Christoffels and De Groot (2005) commented that while transcoding at lexical level occurs, it is likely that transcoded items are at the same time processed further up to full comprehension, leading to a strategy in which both parallel and sequential translation occur.

This model would require more research to afford better resolution in distinguishing the two translation strategies. Having said that, Christoffels and De Groot's (2005) comment should suffice to clarify for readers as to how it is possible that SL input could flow in a translation system in a pure parallel fashion, implying that word recognition can take place at all levels at the same time. Ultimately, the two translation routes this model presents might include one strictly serial processing (sequential translation) and a hybrid type of processing (sequential translation in principle with an option for parallel processing at different levels along the way up and down). With currently available evidence in bilingual lexical access and sentence reading for translation (Macizo & Bajo, 2006; Ruiz et al., 2008) in mind, Paradis's (1994a) adapted model should be able to take bilingual competence into consideration when variables concerning low-level language processing are brought to bear on interpreters' performance. This hybrid type of processing in translation also coincided with the theme *components of translation* I outlined in section 2.2.1 which aimed to address the question of exactly what is left highly activated in translators' memory after parsing a sentence and whether activation of both languages of the bilinguals came with a cost.

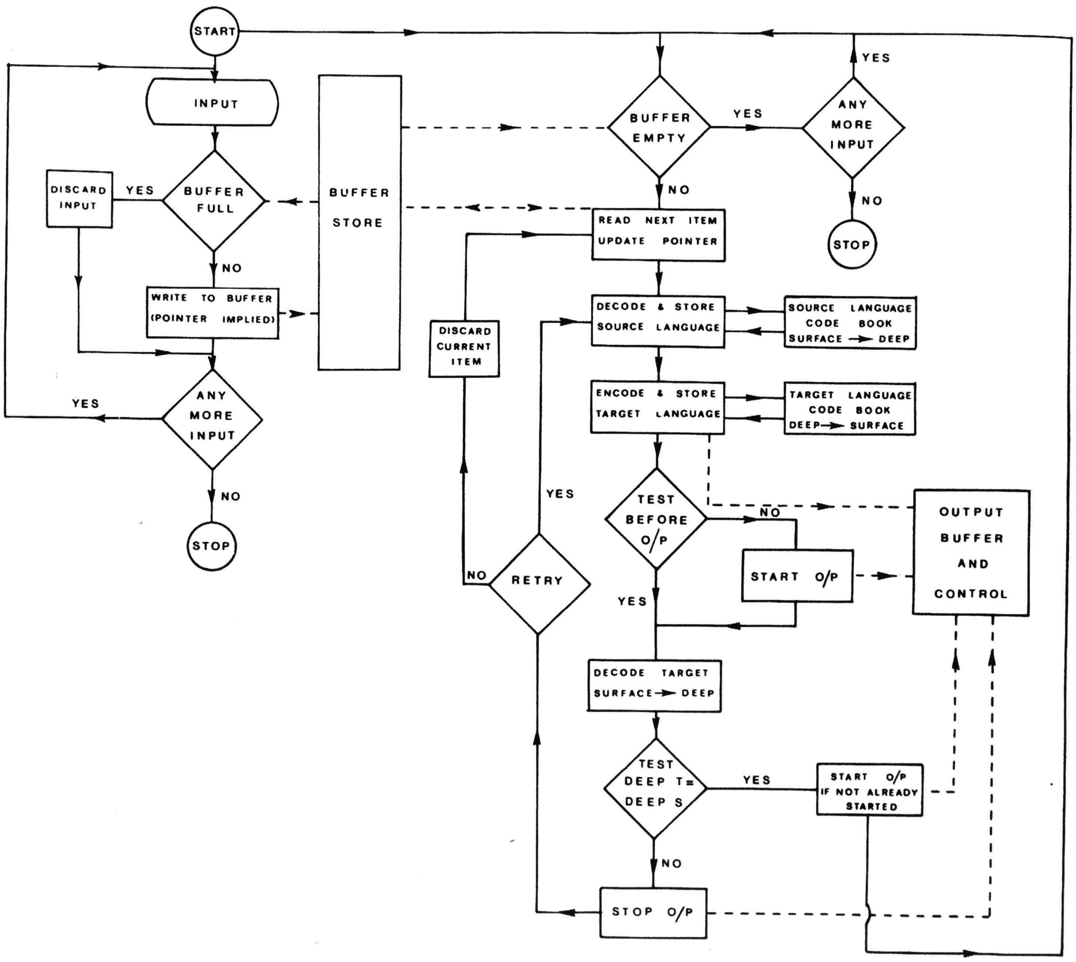


Figure 2.10: A model of simultaneous interpreting, adapted from Gerver (1976).

Around the same period of time when Seleskovitch (1976) put forward her *Théorie du sens*, psychologists David Gerver and Barbara Moser-Mercer responded to the call for empirical approaches to interpreting research by sketching comprehensive models of simultaneous interpreting. Figure 2.10 shows Gerver's (1976) model that was developed on the basis of his experimental findings of participants' ear-voice span (EVS – the stretch of time interpreters' TL production is behind the speakers), memory use and output monitoring (Pöchhacker, 2004b). An important feature of Gerver's model is his postulation about the role of coordination of short-term and long-term memory systems in simultaneous interpreting: "through a process of active reinstatment, linguistic knowledge in the interpreter's long-term memory becomes available in a short-term 'operational memory' or 'working memory' which serves the processing operations involved in source-language decoding and target language encoding. Maintenance in short-term memory is also a prerequisite for monitoring and self-correction procedures which Gerver views as integral parts of the process and particularly vulnerable to temporary shortages of processing capacity"(Pöchhacker, 2004a, p. 100).

More specific to the aspects of language comprehension and production in his simultaneous interpreting model is his assumption of the input and output buffers in order to account for interpreters' ability to translate while more SL material continues entering their language faculty. This assumption is consistent with the idea put forward by Van Hoof (1962) and Hromosová (1972), "as the source language starts, the interpreter begins to store, as he stores he also brings into short-term memory his knowledge of the vocabulary and grammar of both source and target languages, then while pronouncing his translation, the cycle continues" (Gerver, 1976), also see Van Hoof, 1962 and Hromosová, 1972). Another assumption pertinent to the semantic representation of decoded SL material is his distinction between surface form and underlying structure in relation to the discourse context. Gerver (1976) must have been aware of the criticism drawn to Nida's (1969) approach and cautioned that his notion of deep and surface structures may be better understood as "sounds, words, and sentences heard and uttered by the interpreter, and to his understanding of their meaning... these terms do not necessarily imply any particular generative transformational theory of language."(Gerver, 1976) He supposed that translating an English sentence *John has been given a book by Paul* became ungrammatical if it was translated word-for-word into French, the reason being that the rules governing passive structure construction are incompatible between English and French.

Two problems pointed out by De Bot (2000) are the plausibility of the procedure with which input is discarded from the buffer and the underspecification of the term 'item' in his flowchart. Nevertheless, some characteristics of Gerver's model are visionary. It is the first simultaneous interpreting model that has a monitor mechanism that was thought to test whether the meaning materialised in a TL plan, temporarily stored in the output buffer, matches the 'deep' structure of SL text before and during the utterance of translation (Gerver, 1976, p. 199). Gerver (1976) drew an analogy between interpreters' correction and revision in his data and the hesitations and restart in the flow of speech described by Maclay and Osgood (1959). But when is the monitor mechanism put to work? It seems that this mechanism is constantly active, as he suggested that the process of translation also involves the synthesis and analysis of possible translations of the source language message, that is, continuous generation, monitoring, and testing of the translation against the source language message as understood by the interpreter (Gerver, 1976). He also suggested that analysis and subsequent modification (revise or discard) involved word and phrase levels, and it is the interpreters' criteria for adequate performance that determined whether or not a revision/correction must be executed. It is almost impossible to know whether Gerver (1976) considered interpreters' failure in lexical retrieval as one of the causes for phrase revision, but his specification of the model seems flexible enough to account for performance variability due to problems stemming from lexical knowledge. The most interesting aspect of Gerver's model must be the association between the output buffer and TL encoding. Given that the output buffer in Gerver's model was designated a function in temporarily storing the translation output for articulation and monitoring once articulation begins, it would require mechanisms that can loop the representation in the buffer. To the extent that this reasoning is plausible, Gerver's model is also the first simultaneous interpreting model that conceived and implemented an output buffer similar to that in the phonological loop sketched by Baddeley (2003) and discussed by Jacquemot and Scott (2006).

Moser-Mercer (1978) developed a sophisticated model of simultaneous interpreting, based on Massaro's (1975) comprehension theory. In Figure 2.11, boxes represent structural components that describe the product of a previous process; each heading represents a functional component that denotes the type of processing at any given stage; diamonds depict the instances where a decision must be made. Working memory was assumed to be involved in each process in the central column. The coordination between working memory and long-term memory would lead an interpreter to construct a prelinguistic meaning structure, which will activate TL elements to prepare utterances. Clearly, this model also assumes a monitor mechanism responsible for decision making and for looping routines, so that a 'no' response could invoke a process to start all over again. Although she laid out sophisticated stages of word and phrase perception (Moser-Mercer, 2002), it is not clear what happens when access to the SL lexical knowledge in long-term memory is not successful. The same question can be asked of the procedures in language production as well. Theoretically critical questions are whether or not TL production is monitored by the interpreters and whether the demarcation of language systems in long-term memory was backed up by any evidence, since bilingualism studies have shown shared features at different levels between languages. This model is therefore a type of didactic model more than a model ready for empirical testing. As Moser-Mercer (1978) acknowledged, "it has been used fairly extensively for pedagogical purposes and provided the basis for a long-term study on aptitude testing."

Apart from the full-scale processing models introduced above, there are other models that were purported to account for general principles, e.g., Gile's (1999) Effort

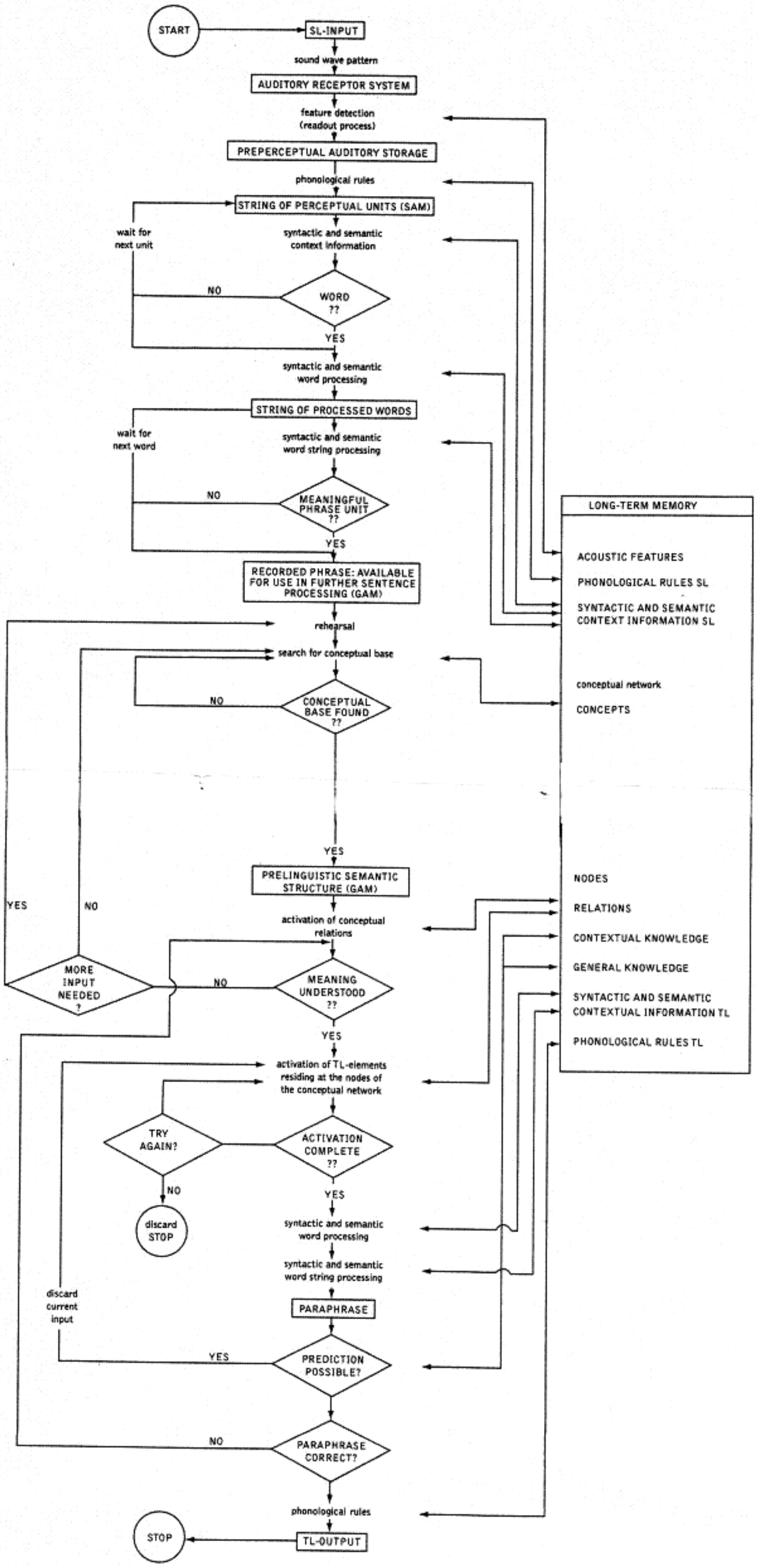


Figure 2.11: A model of simultaneous interpreting, adapted from Moser-Mercer (2002, p. 152).

Model, for *partial* process (Daro & Fabbro, 1994; M. Liu et al., 2004; Mizuno, 2005; P. Padilla, Bajo, Cañas, & Padilla, 1995; Shlesinger, 2000) or for full process of simultaneous interpreting from a perspective of pragmatics (Setton, 1999) (see Moser-Mercer, 2002; Pöchhacker, 2004b for reviews). Although full-scale models have proven their didactic usefulness in training interpreters, they are difficult to be empirically tested. An alternative approach has been to use the full-scale models to guide researchers to formulate specific research questions and use the data to constrain their *partial* models of interpreting. P. Padilla, Bajo, and Padilla (1999) showed how this can be done by giving clear methodological suggestions for particular research questions. The result of applying this principle were a series of studies that have advanced our understanding of language interpretation (Bajo et al., 2000, 2001; Ibáñez et al., 2008; Macizo & Bajo, 2004, 2006, 2009; Ruiz et al., 2008).

In this section, the major translation and interpreting theories were reviewed. Some of these were originally motivated by pedagogical need (Seleskovitch, 1976) and were customised to meet that demand. Many others (Gerver, 1976; Moser-Mercer, 1978) have come very close to theoretically driven models that may become testable. Overall, they do not appear suitable for addressing the questions raised earlier in the section on each translation component, but perhaps they were not designed to do so in the first place. Among them however is an interesting idea (Paradis, 1994a) which has been empirically tested recently (Macizo & Bajo, 2006; Ruiz et al., 2008). The idea that was put to test concerns a possibility of an interaction between bilinguals'/interpreters' two languages at the stage of source language comprehension. By following a similar idea, and recognising the potential of the adapted model (Figure 2.7) in formulating and testing hypotheses of the interpreting processes, this thesis extends Ruiz et al.'s (2008) paradigm to explore the factor of word order and the role that working memory might play in the translation process in a series of experiments.

It is noteworthy that Ruiz et al.'s (2008) paradigm implementation in this thesis was not meant to be an isolated replication study. Because the theories and models reviewed earlier are not suitable in exploring the questions raised for each translation component, a need for theoretical guidance arose and one solution seemed to be adopting established psycholinguistic models as frameworks. The rationale was that without theories of comprehension, memory, and production as framework for formulating research questions, it is difficult to have a full picture of what a translation task involves, and data interpretation would also be difficult.

One prime example of this approach is demonstrated by De Bot (2000) who adapted Levelt's (1989) model of language production (Figure 2.12) in his L2 production model to discuss topics like anticipation, limited attentional resources and language proficiency in simultaneous interpreting.

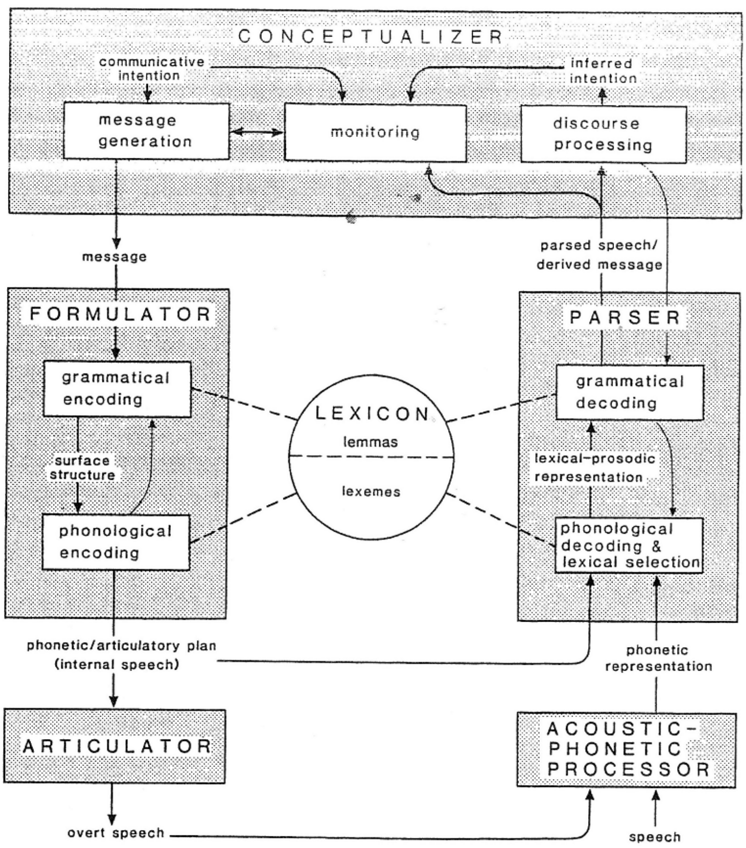


Figure 2.12: A model of language production, adapted from Levelt (1993).

Another example is a bilingual production model in which Kormos (2006) introduced the component of long-term memory in modifying Levelt's (1989) and Levelt, Roelofs, and Meyer's (1999) production models (Figure 2.12). The long-term memory in Kormos's (2006) model (Figure 2.13) consists of episodic memory, semantic memory including the mental lexicon, the syllabary, and a store for declarative knowledge of L2 rules.

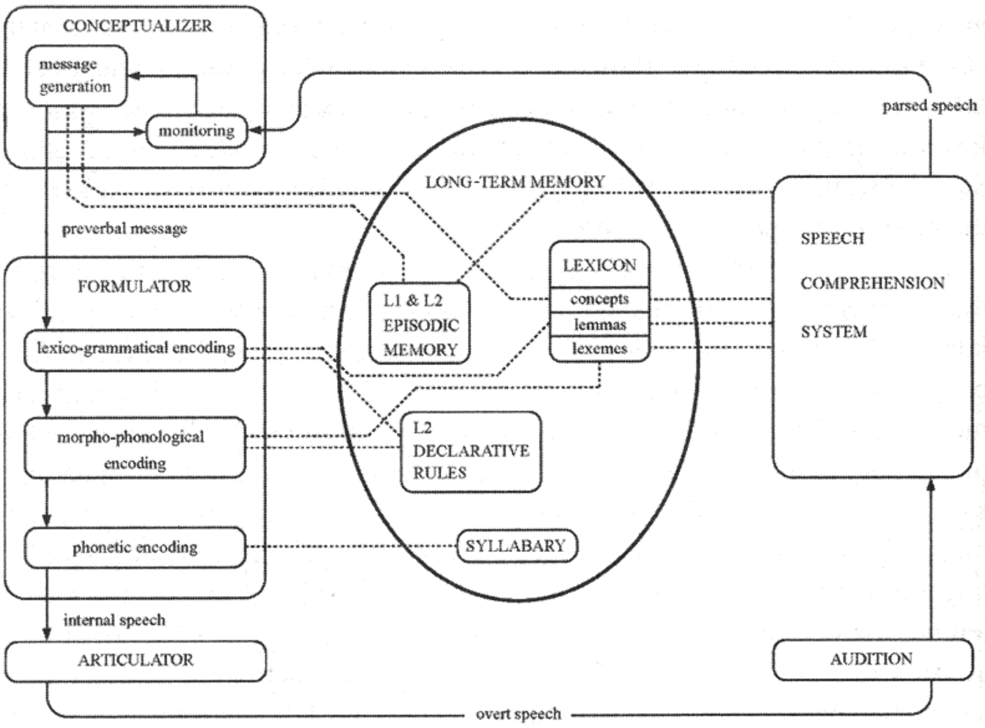


Figure 2.13: A model of L2 production, adapted from Kormos (2006, p. 168).

Because translation involves language comprehension, memory retention[7] and language production, one theory was chosen as a framework to guide the discussion and argument for each of the translation component. In section 2.4, Kintsch and Van Dijk's (1978) Construction-Integration model in comprehension component; Baddeley's (2000) working memory model in resource and memory components; and Levelt's (1989) language production model in grammatical encoding component will be discussed. For the purpose of discussing a goal-oriented task such as

[7] The requirement of memory retention can differ among different types of interpreting. Consecutive interpreting would entail longer retention duration of propositions that appear earlier in a discourse, given that a memory aid is not available.

interpreting and exploring the relation between language encoding and memory retrieval, Levelt's (1989) production model will be used as a general framework. In so doing, it would be possible to relate each of the questions raised in three translation components (i.e., comprehension, memory, and resource) to the component of grammatical encoding. Next section is an overview of Levelt's (1989) model of language production which briefly covers what assumptions have been made when this model is adapted in bilingual language production.

2.2.3 Levelt's model adapted and an overview

Why adapt a production model, instead of a comprehension model? One justification is that a translation task cannot be completed without any form of output (verbal or sign language). In other words, translation is realised through translators' language production. This does not mean comprehension processes are not important. In fact, the assumptions are that bilingual comprehension and production share the same bilingual lexical network (Kroll & De Groot, 1997); and that language production models consist of parameters of language comprehension. This is clearly seen in the parser of Levelt's (1989) model and a postulated mechanism underlying the comprehension-to-production syntactic priming (Pickering & Branigan, 1998) suggests that aspects of comprehension are crucial in production models. Very much like some memory tasks in which recall has to be realised by language production (Acheson & MacDonald, 2009), it is possible to infer the encoding mechanism and the characteristics of the mental representation that support comprehension during translation by analysing the parameters of participants' TL production. By using an appropriate paradigm, it is even possible to investigate participants' comprehension and production at the same time, e.g., Macizo and Bajo (2006). The approach of adopting established psycholinguistic models in bilingual production has some support from D. Green (1993) as he rightly pointed out,

> "There is yet no integrated model of comprehension and production in L2. The adaptation and enhancement of Levelt's model of speech production offers one route to this goal. The advantage of such an approach is that it allows us to state conjectures about the proceses in a way that is not dependent upon the way any model of the process is simulated. Based on such a model, we can view the comprehension and production of L2 as a problem of using resources to control or to regulate the activation levels of various processes and representations."
> (D. Green, 1993, p. 270)

And it is the same approach that has been taken in this thesis to explore interpreting that involves knowledge and skill in more than one language. Below is an overview of Levelt's (1989) model of language production.

The Conceptualiser

Even though the Levelt (1989) model has been modified by incorporating Roelofs's (1992) spreading activation account of lexical access and also by positing a phonetic syllabary which contains articulatory programs for each syllable (Levelt, 1992), the basic principles remain largely the same. The three components in Levelt's (1989) model are the Conceptualiser, the Formulator and the Articulator. To keep the discussion and argument relevant and coherent, the following introduction will be limited to the conceptualisation and grammatical encoding in L1 and L2 language production. As regards the Articulator and phonological encoding, readers are referred to Levelt (1989); Levelt et al. (1999); Roelofs (2003) and Roelofs and Verhoef (2006). The conceptualiser generates intended messages, which involves macroplanning and microplanning. Macroplanning is about the plan of the content whereas microplanning concerns the form of the messages. This process is influenced by what has already been mentioned by the interlocutors in a conversation, (i.e., bookkeeping in Levelt's term), situation knowledge, and the speakers' declarative knowledge. Apart from deciding what to express, assigning focus and topic, and keeping track of exchanged information in conversations, macroplanning also involves ordering information for expression, and this is the process where speakers can have linearisation problems:

> Whenever a speaker wants to express anything more than the simplest assertions, request, declarations, etc., he has to solve what will be called the *linearization problem*: deciding what to say first, what to say next, and so on. (Levelt, 1989, p. 138)

The linearisation problem has been discussed in different discourse domains. Note that the linearisation problem has to be distinguished from the linearisation process in grammatical encoding (e.g., Hartsuiker & Westenberg, 2000, footnote 1), and we come back to this when the Formulator is introduced.) The locus of the linearisation problem in macroplanning was considered to be the preverbal message generation at the stage of conceptualisation in the conceptualiser. According to Levelt (1989), speakers' ordering of information for expression is determined by two sets of factors (see Levelt, 1989, p. 138, for details), but here the focus is the most relevant factor, which is the *natural order*. Levelt (1989) was explicit that there was no general definition of the natural order, but for some domains of discourse, their natural order is obvious. For instance, an event structure's natural order usually corresponds to its chronological order, e.g., *She married and became pregnant.* Another domain is linear spatial structure in such cases where routes or directions are given. A direction giver might think of the final part of the route before working out the initial part in detail. The linearisation problem could occur in other discourse domains such as describing an apartment floor-plan which does not have a clear natural order. For L2 speakers, this problem also applies, but the focus has been placed on the morpho-syntactic level instead of the message level, e.g., observation of sentence-initial adverbials and generation of WH-word sentence in English as a second language. F. Ferreira and Henderson (1998) replicated Levelt and Kelter's (1982) design which required participants to describe networks that vary in their complexity and length in order to test principles of minimal load and incrementalism in message planning in production (Figure 2.14). Their results confirmed both principles: participants preferred short to long branch and they preferred linear to choice branch; participants evaluate the difficulty of two branches rather than the whole network at a decision node. F. Ferreira and Henderson (1998) suggested that "language production system has some tendency to order constituents so that easier material is dealt with before more difficult material... These tendencies can all be taken to reflect a general preference for the most available and accessible word concept to occur early in a sentence."

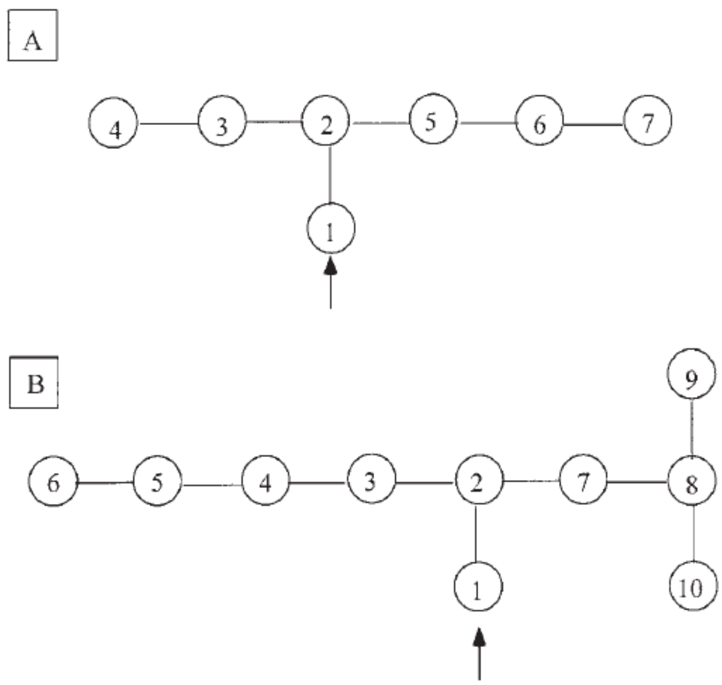


Figure 2.14: Networks used to test principles of minimal load in language production. Adapted from Ferreira and Henderson (1998).

The network description paradigm seems to capture and verify some principles that language production appears to follow. Nevertheless, it shows limitations in its application in language interpreting that is based on the reconstruction of what has been parsed. Given that discourse planning is incremental, an interpreter would not retrieve all he/she can remember before starting to speak. Since a complete *map* of intentions is not available in a strictly incremental language production, it is difficult to conceive as to how linearisation problems come about at the stage of macroplanning in the case of discourse interpreting. But, at least one possibility remains – the linearity problem could be present within a sentence when there are multiple propositions.

The essential requirement of accuracy and fidelity in interpreting could imply that interpreters' translation largely corresponds to the SL discourse structure to ensure coherence, provided that the SL discourse is coherent. If this principle applies to discourse interpreting, an interpreter is not supposed to reverse the order of episodes in a discourse simply because it is easier to retrieve and encode the most recent episode or to start interpreting any episode that is retrieved first from memory. At the sentential level, however, the focus is shifted from episodes to propositions. Multi-propositional sentences can be represented using Van Dijk and Kintsch's (1983) hierarchical structure (Figure 2.15).

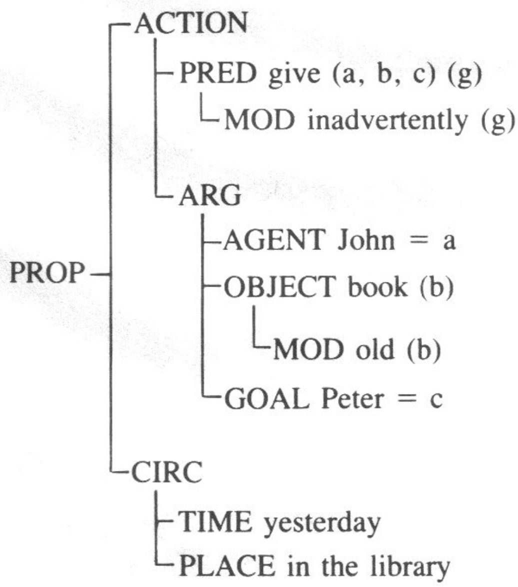


Figure 2.15: A schematic representation of a multi-propositional sentence *Yesterday, John inadvertently gave the old book to Peter in the library.* PROP = proposition; Pred = predicate, Mod = modifier; ARG = argument; CIRC = circumstance category for complimentisers such as time, place, etc. Adapted from van Dijk and Kintsch (1983, p. 115).

The message can be expressed in most languages and represented by similar tree structure in Figure 2.15 but their surface structures will vary to various extents. For instance, sentence 5 would normally be rendered in Chinese word order as:

5. Yesterday, John [in the library] inadvertently gave [Peter] the old book.

Of course one must bear in mind that word order variation often signals topicalisation[8]. But the word order difference in the example above is not a result of topicalisation. Instead, it is a result of different canonical word order between pre-positional modification of verb phrase (VP) in Mandarin Chinese versus post-positional VP modification in English. It would be grammatical if the English word order remains unchanged when it is translated into Chinese. The flexibility of word order in this regard can be a relevant factor because Chinese has more word order variation than English allows. But this word order flexibility, as I have mentioned earlier, has a crucial role in pragmatic connotation (H. C. Chen, 1992). With this kept in mind, when interpreting the example sentence between Chinese and English, the position of VP complementiser *in the library* cannot be arbitrarily assigned, because fronting it to head-initial position, for instance, topicalises the circumstance category of 'place.' As a result, the message in the Chinese translation may be perceived differently from the way it was intended. In order to produce grammatically and pragmatically accurate translation, the word order information has to be somehow encoded during grammatical encoding, probably during the positional process of encoding according to Levelt (1989).

The word concept that first comes into an interpreter's mind cannot always map onto the desirable structure. For instance, in the literature on discourse memory, there is evidence suggesting that atomic propositions can be retrieved more easily than peripheral propositions, e.g., modifiers, because atomic propositions are

[8]For example, topicalising the object 'this' in *I don't know this* can be done by its fronting: *This, I don't know* (see Comrie, 1989 for an overview of typology of language word order)

usually higher in the hierarchical structure of propositions (Kintsch & Van Dijk, 1978). When producing Chinese translation of *John inadvertently gave an old book to Peter in the library,* interpreters might initially encode and produce the atomic proposition (GIVE, BOOK, John, Peter) but realise that they should put the complementizer *in the library* in front of the main clause and consequently restart or correct the translation sentence.

It is possible that in order not to produce whatever comes to mind first and correct the production later, a speaker might have to spend more time in speech planning, especially when a retrieved concept cannot be directly mapped onto a TL surface structure. In this sense, there could be a linearity problem in sentence translation, potentially as a result of structural differences between languages (Levelt & Kelter, 1982; Van Dijk & Kintsch, 1983). This point is elaborated in section 2.4.3.

The stage that follows macroplanning is microplanning, which is thought to assign an accessibility index to each referent in the message, to topicalise referents, and to encode communication intentions to propositional forms, i.e., assignment of perspectives (Levelt, 1989). The product of microplanning is a preverbal message. Since message planning is incremental, "there is no reason to assume that preverbal messages are delivered as integral wholes. Each bit is immediately picked up by the Formulator for grammatical encoding. But the order in which the chunks are delivered will affect the course of grammatical encoding" (Levelt, 1989). The benefit of incremental encoding of preverbal messages on the 'first in, first encoded' basis in the Formulator is the efficient processing that makes speech delivery at speed possible.

The Formulator

According to Levelt (1989), "successive message fragments will trigger the Formulator to access lemmas, to inspect the message for functions, arguments, and modifiers, to specify grammatical relations, and to map these onto inflectional and phrasal structure." Major processes of grammatical encoding in the Formulator were discussed and exemplified in F. Ferreira and Engelhardt (2006) and are depicted in Figure 2.16.

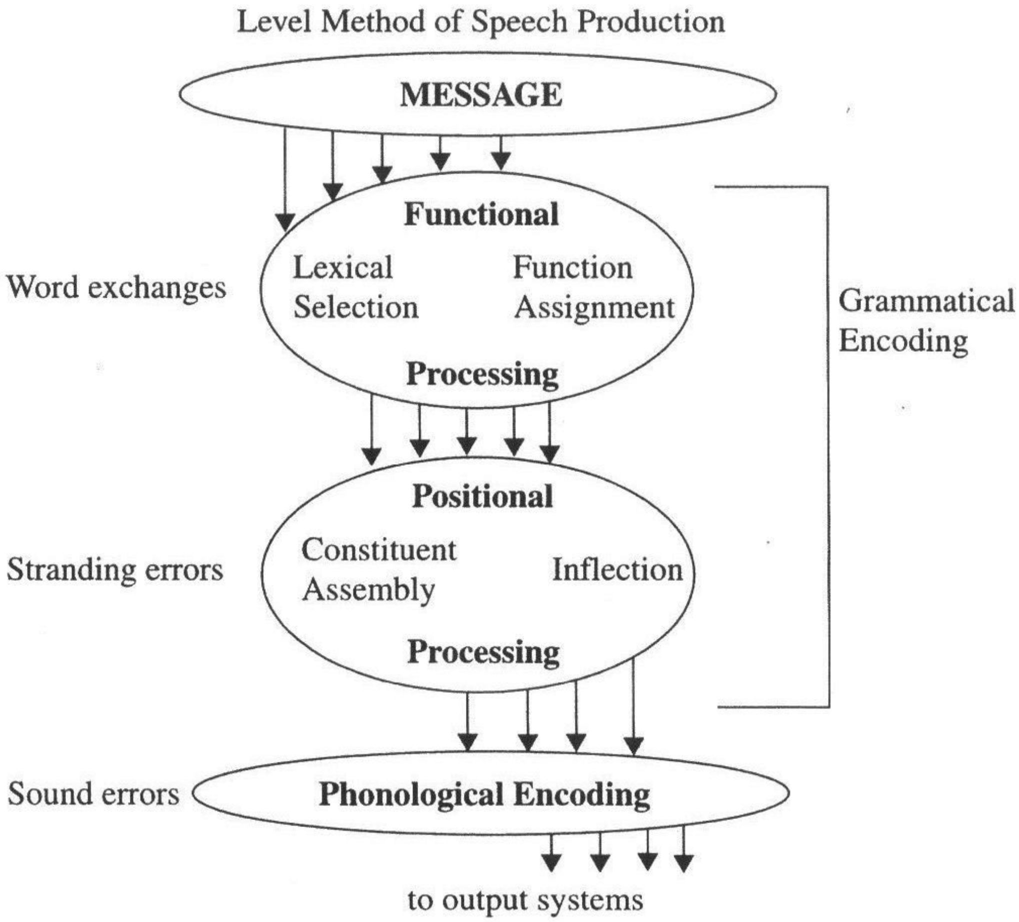


Figure 2.16: An overview of language production. Adapted from Ferreira and Engelhardt (2006).

It is generally agreed that grammatical encoding involves functional and positional levels of processing, but opinions are divided on whether they are taking place in the same stage (e.g., Pickering, Branigan, & McLean, 2002), or in two stages (e.g., Hartsuiker & Westenberg, 2000). Since this debate is not directly relevant to the theme of the thesis, readers are referred to the above- mentioned articles for more details. The discussion here will be focused on lexical access and constituent linearisation with special reference to bilinguals' language production. As mentioned earlier, the modified version of Levelt's production model incorporated Roelofs's (1992) spreading activation model (Figure 2.17) and WEAVER model (Roelofs, 1997).

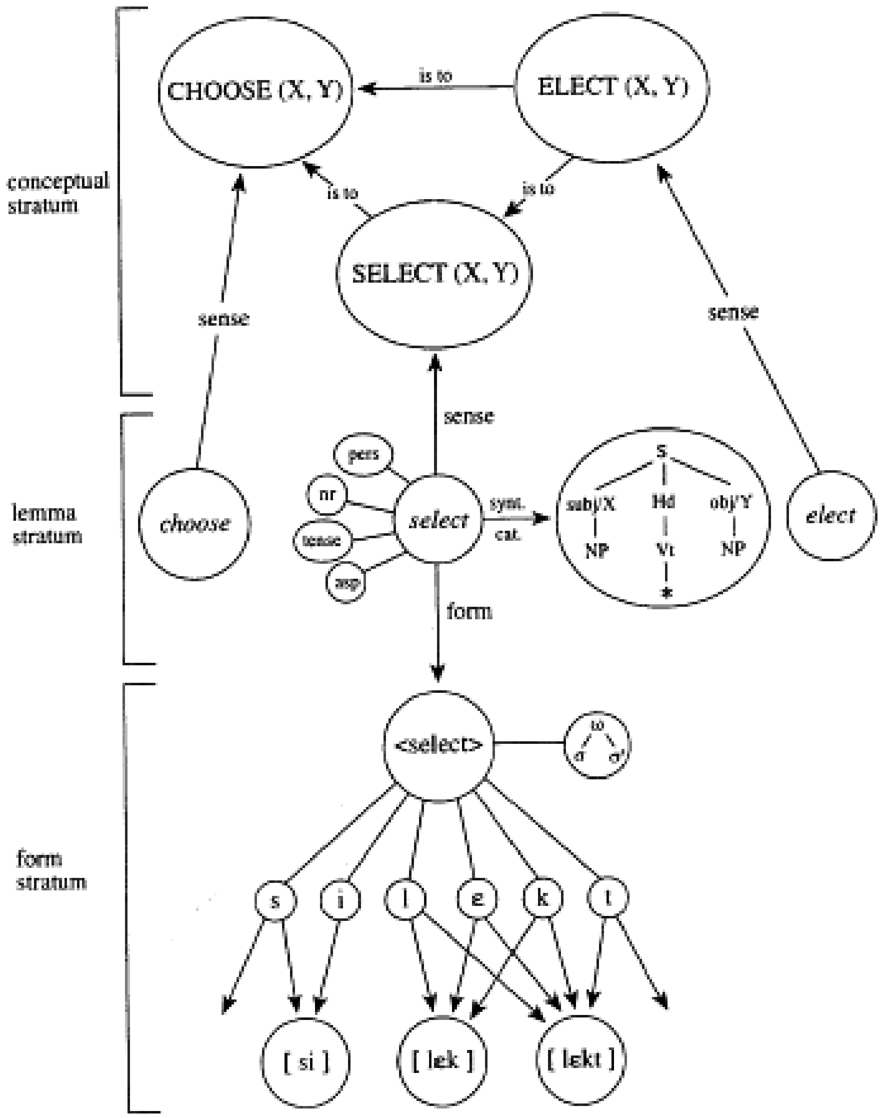


Figure 2.17: Memory representation of the word *select* in lexical network. Adapted from Levelt (1999a).

In the Formulator (Figure 2.12), to-be-expressed preverbal message activates concept nodes at the conceptual stratum, each of which spreads activation further down to the lemma stratum. As Figure 2.17 shows, very often, semantically related lemmas are activated simultaneously. The selection criterion is that the selected lemma must "meet the condition that it entertains the correct sense relation to the conceptual level" (Levelt, 1989). Levelt suggested that the selection process is a statistical mechanism that "favours the highest activated lemma." But it also depends on how 'noisy' the background is: the more semantically related distractors that become activated by the conceptual node, the longer lemma selection takes to complete. Once a lemma is selected, its syntactic information is also activated. For instance, the verb *select* is associated with a few diacritic parameters that have to be specified at this stage, e.g., tense, aspect, etc. The same procedure applies to other lemmas, such as nouns and modifiers. In preparation for producing *my cat terrifies the dog next door*, the lemmas (CAT, DOG, etc) will be retrieved and then assigned functional roles. In this functional level of processing, CAT would be assigned the role of subject, DOG the role of object, which is modified by NEXT DOOR.

Before the discussion moves on to the processing at the positional level, i.e., linearisation, this is the point where bilingual production needs to be reflected upon. Recall that lexical access is thought to be language-nonselective (section 2.3.3), therefore a preverbal message spreads activation not only to lemmas in a TL lexicon, but also those in an SL lexicon. Even when a lemma is selected, bilinguals showed influence of SL on the pronunciation of TL words (Poulisse, 1999). For late bilinguals, SL interference might be trivial compared to a common problem of incomplete L2 lexical knowledge in multiple dimensions, one of which is the relatively smaller productive L2 vocabulary than that of L1. The question concerns what happens when the sought-for lemma is not there. According to Levelt (1989), "the conceptual system is ignorant about the accessibility of lexical items... The present evidence suggests that message preparation proceeds in full ignorance of the developing state of affairs at the levels of grammatical and phonological encoding." In the case of L2 production, L2 speakers either have to regenerate a preverbal message or resort to communication strategies when they encounter lexical retrieval problems (Poulisse, 1997, also see paragraph **Automaticity in lexical and grammatical information access** in section 2.2.2). When a speech plan must have another take, it often results in hesitations or filled pauses between utterances. The problems facing L2 speakers, however, are not limited to the stage of functional process in grammatical encoding. When lexical items are selected, the ease with which these items become sequenced can partially determine the time it takes speakers to initiate speech output (Corley & Scheepers, 2002; F. Ferreira, 1991; Smith & Wheeldon, 2001). It will be shown in the following paragraph that L2 speakers' competence will face another test in how fluently selected lexical items are sequenced in a correct order, in addition to parallel operations up and down the stream (e.g., conceptualisation and phonological encoding).

In Levelt's model, another stage of processing that is thought to take place either in the same stage of function assignment or in a following stage is positional processing or linearisation. At this stage, serial order is imposed on the utterance. Kempen and Hoenkamp (1987) postulated encoding procedures for major category lemmas, e.g., PP- procedure for building prepositional phrase. Each selected lemma calls specialised syntactic procedures to produce a surface structure for the next stage of encoding. But how is the order decided? For an utterance like *my cat terrifies the dog next door*, there is no a priori reason why it must be *cat* that appears in the clause-initial position. "The left-to-right ordering is merely a notational convention. The order in which the message fragments become available may be a different one. That order can be a major determinant of the eventual syntactic form" (Levelt, 1989, p. 237). Levelt went on to suggest that "Fragments of a message are grammatically expressed as much as possible in the order in which they became available... topical or salient concepts tend to be expressed early in the sentence" (Levelt, 1989, p. 239). F. Ferreira and Engelhardt (2006) also commented "because the language production system is assumed to be incremental, the order in which lemmas are 'worked on' determines the overall order of the phrases in the utterance... So if the lemma for DOG were processed before the one for CAT, then the resulting structure might be the left-dislocation form given above [*the dog next door, my cat terrifies*]." Another possibility for DOG to be encoded earlier is when it is topicalised, and consequently a speaker is more likely to use passive construction: *the dog next door is terrified by my cat.*

A general principle called "conceptual accessibility" is proposed by J. K. Bock and Warren (1985) who showed that a highly accessible concept tends to be encoded in a prominent position and appears early in an utterance. Of direct relevance to discourse interpreting are studies that used probed sentence recall (J. K. Bock, 1977; J. K. Bock & Irwin, 1980). In J. K. Bock (1977), participants were presented with a list of sentences and then they were given a question as recall cue for one of the sentences on the list. One study sentence was *a psychologist cured a neurotic poodle.* J. K. Bock (1977) constructed the probe question such that it topicalised one particular entity in the to-be-recalled sentence. The question that followed the study sentence was: *The interior decorator was afraid she would have to get rid of her neurotic pet poodle because it was ruining the furniture, but she was able to keep it after all. What happened?* It turned out that participants were more likely to produce a sentence in the passive voice: *the neurotic poodle was cured by a psychologist.* This finding was taken by J. K. Bock (1977) to suggest that when a topic is assigned by speakers, it tends to be encoded in the sentence-initial position as the subject. In J. K. Bock and Irwin (1980), recall probes were keywords, and they found that participants tended to start a sentence placing the probe in the head position. If a study sentence was *The falling tree crushed the lumberjack*, participants tended to recall it with active voice when the probe was *tree* and with passive voice when it was *lumberjack.* When a translator is retrieving discourse content while grammatically encoding retrieved information, it is very likely that similar principles or mechanisms operative in a way similar to those found in spontaneous speech will be observed, in that the principle of accessibility is reflected in the early appearance of atomic propositions in a translation, followed by propositions of arguments and modifiers, that vary in the degrees of importance in a discourse.

The incrementality in grammatical encoding has complications and implications. Although it has been thought that incremental encoding "minimises the need to buffer constituents and hence facilitates sentence formulation" (Bernolet, Hartsuiker, & Pickering, 2007, p. 932), it "might create a situation in which an accessible constituent forces a syntactic structure that is computationally demanding (e.g., the passive)"(F. Ferreira & Engelhardt, 2006). F. Ferreira and Engelhardt (2006) however suggested that "the difficulty of making a passive can be 'spread out' over the entire utterance rather than being localised entirely to the point of its initiation. As a result, there need not be any hesitation or disfluency before utterance production, and the demands of managing the rest of the structure can be distributed over the remaining constituents, with planning going on in parallel to articulation." One implication of incrementality of encoding pertains to how the process of memory retrieval in interpreting and the process of language production are coordinated by interpreters. The discussion of translation components in section 2.4.3 expands on this point.

For now, consider how interpreters fare in the linearisation process in a translation task. The catchy terminology of linearisation or sequencing tempts a hypothesis that translation and interpreting involve an operation of constituent movement, because the word orders differ between languages most of the time. Attractive idea as it appears, no evidence is currently available to support it. The closest idea (e.g., trace and constituent movement in the Government and Binding theory) that can be found in linguistics and machine translation (H.-H. Chen, 1992), however, has very little to offer to language interpreting, as empirical evidence for the psychological reality of constituent movement is not yet available even in monolingual language production. Currently, from the perspective of Levelt's (1989) model, one objection to the movement hypothesis in interpreting is that it violates the incremental processing of production. For instance, if the movement hypothesis holds, it would require an interpreter to hold and manipulate discourse or sentence propositions on-line until a grammatical TL sentence is encoded in its entirety for output in translating a Mandarin Chinese sentence 我們上個禮拜在公園裡野餐 into English:

6a. Chinese: 我們(we) 上禮拜(last weekend) 在公園裡(in a park) 野餐(went on a picnic)。

6b. English: We went on a picnic in a park last weekend.

This hypothesis runs against the generally accepted principle of the economic use of cognitive resource. Indeed, it could tax a great deal of working memory capacity in order to loop these items, and consequently the system would be overloaded. By following the incrementalism that word order is initially underspecified in the conceptualisation and functional processes, until the order is specified in positional process, an interpreter is likely to start encoding the most salient and accessible concept, e.g., the subject, and then the verb phrase, its modifiers, and so on. The ease with which linearisation is carried out could probably depend on the degree of automatisation of the syntactic procedure of each grammatical category (e.g., NP-, PP-, S- procedures) and also on their similarity between languages, as Prodeau (2005) suggested, "procedures that are similar in both languages have been automatised and thus use very little space in working memory, so there is little risk of memory overload."

When the incremental processing in L1 production allows parallel operation of other processes, i.e., concurrent conceptualisation and phonological encoding, grammatical encoding in L2 has been assumed to be serial: "the lack of automaticity could simply be captured by assuming serial, step-by-step processing (at least to some extent) instead of parallel processing at the morphophonological and articulatory levels. As serial processing is slower, it allows the speaker to replenish the resources needed to carry out non-automatic, attention-demanding processes" (Poulisse, 1997, p. 208). The cost of cognitive resource in the stage of grammatical encoding in L1 has been attested by increased tracking errors around the gap at the beginning of a subject-extracted relative clauses and also at the end of a object-extracted relative clause in a dual-task design, suggesting that a checking mechanism is in place to assign a role to an appropriate referent in working memory when there is a gap in a sequence (Power, 1986). F. Ferreira (1991) manipulated sentential subjects in their length and complexity in a sentence recall task. Sentence initiation time increased with the complexity of the subject, but not the object. When both the subject and the object were complex, participants tended to pause within sentences. It thus appears that grammatical encoding is not entirely automatic, and it might depend on the syntactic complexity of the to-be-uttered sentence. Grammatical encoding is not cost-free, therefore, the finding that speakers' initiation time decreased when their produced sentence structure was similar to the prime sentences was significant. The facilitation of priming on speech initiation (Corley & Scheepers, 2002; Smith & Wheeldon, 2001) suggests that syntactic priming could be a tool used by the production system to "reduce the load associated with syntactic processing" (Pickering & Branigan, 1999).

Taken together, current evidence and production theory should allow predictions in translation as follows. When a syntactic structure or word order is shared by languages, it is likely to be reused in the TL utterance, and one indicator of such a priming effect across languages is the initiation time. But when there is no clear indication that structures can be reused, interpreters' production performance, similar to L1 speakers, depends to some extent on the task requirement. If the task encourages TL production as soon as possible, the production system might become more incremental and more efficient, i.e., the same amount of work can be done in a shorter amount of time than when there is no time pressure (F. Ferreira & Swets, 2002). In which case, it is unlikely that there would be any difference of speech onset time in TL output in translation irrespective of discourse characteristics, because participants have to start speaking immediately after the presentation of the SL discourse is completed.

According to F. Ferreira and Engelhardt (2006), the difficulty associated with encoding can be 'spread out' over the entire utterance, thus it is likely that early onset of speech output entails a less-than-fluent process in the subsequent incremental encoding. But since the cost can be 'spread out', the debt is smaller and less localised (F. Ferreira & Engelhardt, 2006), and this cost may be captured by the average speech rate and pause probability/duration of the utterances. There is evidence that L1 speakers appeared to prefer planning more carefully before speaking (V. S. Ferreira & Pashler, 2002). It is likely that when a task affords L2 speakers to plan, participants would take advantage of it. In contrast to relatively earlier onset of speech output under incremental processing, planning ahead of utterance might take time and postpone speech onset, but the result of planning could be an utterance with higher speech rate and/or fewer hesitations.

In the brief review of Levelt's (1989) model, the scope of discussion was limited within the conceptualisation and grammatical encoding processes and raised a key question: how do interpreters coordinate memory retrieval for discourse content while preparing each retrieved fragment of discourse for language production in a TL? This question involves issues that are beyond the scope of this thesis. For instance, in the context of language interpreting of passage-long discourse without memory aid, is there a need to distinguish memory retrieval for to-be-translated content from conceptualisation in TL production? To discuss a more directly relevant question, if retrieving fragments of discourse content is isomorphous to conceptualisation, the focus can be placed on the interaction between the memory of parsed SL discourse and the way in which it is linearised in TL production. Consider a paragraph as follows,

> Pain has a strong psychological component. What the patient is thinking can determine his or her pain scale. Human attention allows us to select some information to process and to ignore everything else. Shifting patients' attention by, for example, having them listen to music can help reduce pain.

In order to interpret this passage, an interpreter would normally take note to aid his/her memory. But in cases when note-taking is not possible, an interpreter would have to somehow memorise the whole passage. Some interesting questions then can be asked when memory retrieval and language production seem to occur at the same time. First of all, one may question how the prose is memorised: is the representation of the prose more like a network of connected nodes from a connectionistic approach (Kintsch & Van Dijk, 1978) or it is more compatible with chunks of information buffered in working memory (Baddeley, Hitch, & Allen, 2009)? Since L1 speakers can have linearisation problems at the stage of conceptualisation and L2 speakers could have problems in the linearisation process in grammatical encoding, it is worth also exploring whether the linearisation issues affect interpreters' performances in terms of their memory recall for content, e.g., proportion of reproduction, and the measures of language production, e.g., duration of speech initiation.

Assumptions made in bilingual production models

Extra assumptions have been made in adapting a language production model for special purposes, just like De Bot (2000) and Kormos (2006) did for their models. Four assumptions made in De Bot's (2000) model (see De Bot, 1992 and De Bot & Schreuder, 1993 for details) are as follows:

- All linguistic elements have to be labelled for language. In particular, the preverbal message has to specify what the language is of each chunk it conveys to the formulator.

- There are stores for lexical items, syntactical rules, morphemes, and syllables. Within these stores there are language specific subsets that develop over time through use (following proposals by Paradis (1987).
- The activation spreading is the main mechanism in the selection of elements and rules. Elements and rules differ in level of activation, and the ease of retrieval is dependent on this level of activation: high frequency words can be accessed more easily and faster than low frequency words.
- As languages can be accessed as subsets on different levels, they can be more or less activated as a whole, and languages will have differing level of activation depending on level of proficiency, setting and recency of use (De Bot, 2000).

Although De Bot (2000) used his L2 production model in accounting for phenomena in SI, he did not limit the application of his model to SI only. The factors that were deemed crucial in simultaneous interpreting may also need to be taken into consideration in discussing translation in other different forms, since the demand from different forms of translation has been suggested by Campbell and Wakim (2007) as any point on a continuum between 'slow translation' and 'fast interpreting' modes, depending on time constraint. According to the time allowed to translators in executing their production in translation, Figure 2.18 shows that the key parameters in translation vary in terms of the frequency in attention switching, speed required for transcoding, probability of preserving original word order and the amount of ST surface structure retrievable in memory.

In this line of reasoning, sentence consecutive interpreting in which translators alternate between SL comprehension and TL production in fast succession would fall on the continuum near the fast end, whereas consecutive interpreting for a long passage would fall near the slow end of the spectrum. In both cases, translators are expected to start delivering translation as soon as a chunk of SL discourse is

		Production modes				
		Slow translation	Fast translation	Sight translation	Consecutive interpreting	Simultaneous interpreting
Elements of mental representation	Attention switching	slow ←				→ fast
	Automaticity as transcoding	low ←				→ high
	Potential for word order variation	high ←				→ low
	Retention/decay of ST formal structure	high ←				→ low

Figure 2.18: Translation-Interpreting Continuum, adapted from Campbell and Wakim (2007).

finished, therefore, the time pressure in long passage interpreting is not much different from that in sentence interpreting. De Bot (2000) did not put his model to test, but he reviewed articles to illustrate aspects in interpreting that a model must be able to address. Anderson (1994) compared between-language interpreting with within-language shadowing (text repetition) to address the question as to whether translation took an extra stage than within-language interpretation. Bilingual participants' EVS (Ear-Voice Span) measure was longer in interpreting condition than in repetition condition, but there was no difference in directionality within interpreting or shadowing tasks. Anderson (1994) interpreted this result as an evidence for an additional translation stage. De Bot (2000) used this case to show that under the assumption of limited attentional resource available to interpreters, attention allocation will depend on how automatic interpreters' subskills are, e.g., for word retrieval and syntactic encoding. If lower-level processing was not automated but required controlled processes, this will slow down the higher-order processes, e.g., conceptualisation or macroplanning which are assumed to be the most demanding by language production models. When this argument mainly applied to the production phase of interpreting, slower processing under constraint could apply to the comprehension phase equally well. As sentence comprehension involves multi-level interactive processing, comprehenders' working memory has been implicated in holding candidate interpretations for ambiguous lexical items or syntactic constructions (MacDonald, Just, & Carpenter, 1992; Miyake, Just, & Carpenter, 1994). Working memory has also been associated with individual's ability in rapidly integrating information where higher-order computation is required, e.g., pragmatic cues (Carpenter, Miyake, & Just, 1995). Fuller elaboration of the role of resource allocation is in the discussion of comoponents of comprehension, memory and grammatical encoding. Specific questions are considered in the section 2.4.4.

In the same vein of reasoning, De Bot (2000) went on to associate language proficiency with the timing aspects of translation process. On the basis of Driessen's (1993) data of bilingual word translation, he established effects of translation directionality (similar to the translation asymmetry reviewed in section 2.3.3) and language proficiency, such that translation latencies were shorter for the more proficient speakers. The most interesting result was reported in a word pair verification task, in which Dutch-French bilinguals had to indicate whether a pair of Dutch and French words were translation equivalents of each other. Although there was a difference in reaction times between two of the three groups of participants who were divided by their proficiency, their accuracy in verification suggested that "the lexical knowledge about the words tested is still present, but producing the words under time pressure, especially in the L2, remains a problem, even for the advanced learners" (De Bot, 2000). This observation is consistent with late bilinguals' larger passive (receptive) than active (productive) L2 vocabulary (see section 2.3.3). A conflict was spotted and questioned by De Bot in the effect of directionality between the better performance in backward (L2-L1) word translation and the relatively better passage translation into interpreters' L2 (Barik, 1975). Note that the latter phenomenon is contradictory to a widely applied routine that interpreters translate into their L1 and a review of literature indicated that no studies to this date had replicated Barik's result of better translation quality in L1-L2 than L2-L1 interpreting. De Bot's (2000) comments, therefore, warrant further investigations and this thesis could contribute to partially address the paradoxical observation:

> "... it is clear that there is a gap to be filled: the fact that word translation is faster from the weaker language into the dominant language than from the dominant language into the weaker language is an information processing fact. The other fact, that interpreters are more successful when translating from their dominant into their weaker language than the other way around, shows that choice of strategies has a larger impact on the whole process of interpreting than the retrievability of words from the lexicon. In other words: even though it takes more time to find the right words, the advantage of a better and deeper understanding of the incoming speech more than compensates for this." (De Bot, 2000, p. 85)

While De Bot (2000) was not very explicit about the predictions that can be made for interpreting performance on the basis of his assumptions, Kormos's (2006) model that was adapted from Levelt's (1989) model drew studies in second language acquisition, language production, and bilingualism to elaborate on how her model can account for the use of communicative strategies, code-switching, phenomena of language transfer, and the way in which formulaic language and encoding procedures were developed. In order to achieve that, Kormos (2006) introduced long-term memory in modifying Levelt's (1989) model. The rationale of this modification seems to be related to the developmental aspect that Kormos wanted to highlight from the perspective of second language acquisition. The oval structure in Figure 2.13 was introduced to replace the lexicon in Levelt's model and to illustrate components that were crucial to L2 production. A designated store for syntactic and phonological rules of L2 was thought instrumental in Kormos's model to reflect several differences between L1 and L2 production: 1) L1 tends to influence L2 processing; 2) L2 knowledge is often incomplete; 3) encoding in L2 is often not automatised and requires attentional resources. Kormos suggested that because of these developmental characteristics, L2 speakers are prone to encounter problems or are sensitive to issues like resource deficits (e.g., lack of lexical knowledge), time pressure, perceived deficiencies in one's own language output, and limited attentional resources. With these specifications and assumptions taken into consideration, this model seems to have most of the important aspects that bilingualism research has shown in characterising bilinguals (see sections 2.2.2, 2.3.3) and it also can respond to Campbell and Wakim's (2007) methodological questions about translation studies. Campbell and Wakim (2007) argued that the prerequisites for a model of translation task must have characteristics as follows:

- The model must include the switching of attention between comprehension and production.
- It must make reference to automaticity in some way in order to account for the difference between low speed and high speed tasks.
- It must deal with the issue of word order, i.e., the variation across production modes, in their potential to present target text information in a different order from the ST. This issue is related to the "linearisation problem" (Levelt, 1989).
- It must allow retention or decay of the formal structure of the source text, e.g., in fast translation and consecutive interpreting respectively.
- It must include the developmental dimension to account for production in the second language. This dimension is theoretically justified in line with Campbell's (1998) view that translators into L2 are by definition learners of L2 (Campbell & Wakim, 2007).

Apart from the first characteristic, Kormos's (2006) model appears qualified for discussing the other topics listed above. And crucially, these topics are central to the research questions in this thesis. In order to frame and address the research questions, Kormos's (2006) model was chosen as a framework, as it is based on an extensively studied model of language production by Levelt (1989), and its developmental dimension incorporates the robust findings in bilingual acquisition and competence. An overview of Levelt's (1989) model has already been given, and now the focus shifts to an overview of the important developmental dimension highlighted in Kormos's (2006) model. The next few sections review relevant bilingualism research that should help address the questions 1) can bilinguals translate? 2) and if they can, what type of bilinguals are these experiment participants from a developmental point of view? 3) how have these type of bilinguals been characterised in psycholinguistic studies and what are the implications of these characteristics in doing a task like language interpreting?

2.3 Some facts about translators

A long-standing methodological issue in psychological research on translation and interpreting has been the underspecified linguistic profile of the participants. Unlike the research on monolingual comprehension and production studies, an inevitable but often criticised assumption that has been made in bilingual or translation studies, is the bilinguality of the participants. This kind of assumption can be counterproductive because the impact of the cognitive demand in translation, which can be captured by the four components in the previous section and manipulated in an experiment, may not speak to theories or hypotheses of interest if participants' linguistic competence of languages is not taken into consideration by experimenters and accounted for by models or hypotheses in focus. Since "what enables translators to cope with these tasks [variable tasks that make specific demands on the cognitive system of the translator] is their translational competence," (Neubert, 2000) it is essential to profile translators' competence to enable a coherent and viable interpretation of translation performance. This section sets out to characterise bilingual experiment participants in three respects: 1) are bilinguals natural translators; 2) translators' bilingual acquisition; and 3) just how bilingual they are. It will conclude by considering the implication of participants' linguistic competence and bilinguality when translation is discussed in models of translation, language production, and working memory respectively.

2.3.1 Are bilinguals naturally competent translators?

Translators take up the job from different backgrounds. An interpreter working for the Taliban's ambassador to Pakistan was an Afghan fighter during the Soviet invasion and later did business in the Bayside section of Queens, N. Y. for several years (Amos, 2009). On the battlefield, translators could be any local who may be fluent in English and any dialect, such as *Pashto* (Synovitz, 2008). Young bilinguals in immigrant families have been known to act as family interpreters in an occasion like medical examination from as young as eight years of age (Kaur & Mills, 1993; Malakoff & Hakuta, 1991; Orellana, Dorner, & Pulido, 2003). The population of these go-betweens who have been dubbed 'information brokers' or 'language brokers' should easily outnumber the so-called 'translators' under a narrower definition of translator. However, this population has attracted little attention in the area of bilingualism or translation research, even if their skill is relevant from both socio-cultural (Toury, 1995) and skilled behaviour point of view. This skill is called *natural translation* by Harris (1977; 1978) who has a background of machine translation. B. Harris and Sherwood (1978) suggested that 'natural linguistic skills' enable natural translation, and therefore, natural translation is a necessary concomitant to bilingualism. Second, translation skill can improve over time under guidance, just like any other natural skill. Natural translation as bilinguals' innate ability was generally supported by observations of translation by young children when the to-be-translated materials were within their comprehension and vocabulary (Hakuta & Malakoff, 1987; Hakuta, Gould, Malakoff, Rivera, & Rodriguez-Landsberg, 1988; B. Harris, 1980). Although natural translation has not been brought up in the literature of translation and interpreting studies that focus on the cognitive processes, a small survey at the end of section 2.3.2 reveals that the assumption of natural translation as innate ability of bilinguals may have been made by researchers implicitly, and more widely, in bilingualism and translation studies than it seems. The implications of this assumption can be significant. One is that interpreting the results of experiments using *natural translators* may be easier than when participants were trained translators, because the lack of formal training implies less influence of coping skills and strategies on their processing during translation, i.e., data is less contaminated. A similar argument has been put forward by Barik (1972).

Barik was able to support his argument with the finding of a narrow margin between amateur and professional translators' performance. If Barik's idea holds, recruiting natural translators in translation studies may be justified, but the downside is that one cannot be sure how 'natural' they are. As B. Harris (1977) put it, "in educated communities it is only by catching translation at a very early age that we can be sure of observing it in its 'natural' state." Natural translation can be improved over time, but with increasing practice, "the act of translating loses its naturalness (Kaya, 2007). If this naturalness is only present before formal language education or training, maybe the natural ability of translating is only one of the stages of becoming a bilingual. A more important determining factor than the naturalness of translation act in the development of translation competence would appear to be the context of bilingual acquisition. This consideration incorporates the notion of translation competence which highlights a multi-dimensional competence including linguistic, cultural, and strategy competence (Bell, 1991; Hewson & Martin, 1991; Neubert, 2000; Gideon, Anthony, Miriam, & Daniel, 2008). The common requirement for translators among translation competence theories is the "near perfect knowledge of the grammatical and lexical system" (Neubert, 2000). To avoid overlooking the significance of other parameters than linguistic competence, one way of determining how competent a bilingual is in translation is by examining the context in which his/her languages were acquired, which has to be complemented by qualifying a bilingual's linguistic competence. This is because different contexts of language acquisition offer different linguistic and socio-cultural input in both width and depth. The lack of either thorough understanding of cultures or command of languages would make translation a tough proposition, especially for bilinguals with very basic bi-directional communication skill (Presas, 2000). Following this line of reasoning, given that natural translation is coexistent with bilinguality, the question of whether bilinguals are competent translators will be addressed in a two-stage analysis by determining bilingual individuals' type of bilinguality in section 2.3.2 and then profiling their linguistic competence in section 2.3.3.

2.3.2 Bilinguals' status of bilinguality

The inquiry in the first stage of analysis is one of when and how languages were acquired in the first place. Baetens-Beardsmore's distinction between compound and coordinate bilingualism was based on the difference between a language acquired vs. a language learned, "the compound bilingual having acquired two L1s, the coordinate having acquired one L1 and learnt an L2 at some stage beyond the critical age of more or less 11 years" (Baetens-Beardsmore & Lee, 1975). With a clear emphasis on the temporal sequence and the critical period of language acquisition, McLaughlin (1978) made a distinction between simultaneous and successive acquisition of languages. Bilinguals can also be divided by their language dominance into balanced and unbalanced bilinguals. The dichotomisation of bilinguals, however, is only relevant when their predictions, if any, are attested by the performance of bilinguals from different categories. In this section, the literature review of bilingualism studies will reveal the performance difference between bilinguals whose languages were acquired in different context, and it will be discussed whether ultimate attainment of a second language implies that a bilingual is a native in his/her two languages. The discussion will conclude by considering the implications of incomplete acquisition of bilinguals' second language when translation as a natural skill of bilinguals is placed in models of translation and language production.

"The practice of alternately using two languages will be called *Bilingualism*, and the persons involved *Bilinguals*" (Weinreich, 1953). By this definition, the Taliban interpreter, the Afghan villager, and the young immigrant in the previous section can all be called bilinguals. But when bilingual competence cannot keep up with more demanding translation, translators' loss in translation has often lead to fatal mistakes in a battlefield (Synovitz, 2008). A worry of a less life-threatening scale than translation mistakes made in a battlefield, however, has an equally far-reaching effect in the field of bilingualism research. This is because bilingualism research, with few exceptions (Kroll & Tokowicz, 2005), has shown a tendency to overlook the factors related to experiment participants' history of bilingual acquisition and their bilingual status. Also, in psycholinguistics, bilinguals are often regarded as language users rather than as translators (Presas, 2000). Nevertheless these bilinguals take part in studies that require translation. When these bilinguals' performance is used to develop theories of bilingualism and translation, its viability and generalisability is questionable. In the light of the fact that research that uses translation tasks in bilingualism or translation studies often relies considerably on bilingualism theories, a psycholinguistic profile of the bilingual participants is warranted. It can enable a better understanding of the innateness of translation in bilingual population. It may also provide a more solid theoretical ground for the generalisation of the findings of bilingualism research, whose participants often share the linguistic profiles of those in translation studies.

In the early literature of bilingualism, a plethora of taxonomies for types of bilingualism (coordinate vs. compound; simultaneous vs. successive; early vs. late;

balanced vs. unbalanced) seemed to have contributed as much intellectual discussion as confusion. In Kroll and Tokowicz (2005, p. 542), there is a unification of the two:

> "Generally defined, *compound* bilinguals are individuals who learn two languages *simultaneously*, in the same context, whereas *coordinate* bilinguals are individuals who learn their two languages in *succession*, in separate contexts."

According to W. E. Lambert (1963), learning two languages within one context results in a compound system in which the symbols of both languages function as interchangeable alternatives with essentially the same meanings. On the contrary, learning languages in a context that is "culturally, temporally, or functionally segregated" leads to a coordinate system whose two sets of symbols would be more distinct and independent, with a one-to-one correspondence between each word and its meaning. The difference of learning context was shown to be the capability of distinguishing the meaning of the same word across languages (Jakobovits & Lambert, 1961; W. E. Lambert, Havelka, & Crosby, 1958) and the selective impairment to bilinguals' languages when they became aphasic (W. E. Lambert & Fillenbaum, 1959). This distinction between compound vs. coordinate bilingualism was suggested to be useful in demonstrating how second-language words were context-specifically encoded, and in the study of bilingual autobiographical memory because episodic memory could have language-specific cues (Heredia & Brown, 2004). However the notion of compound-coordinate bilingualism is less influential when a discussion is about language competence and attainment.

Another dichotomy of bilingualism concerns a century-old debate on the critical period that distinguishes early vs. late bilinguals. The reason why age is important is a hypothesis called native language neural commitment (NLNC) that links early language experience to language learning later in life. Kuhl (2004) argued that dedicated neural networks form for languages that are learned early in an individual's life. These networks promote future use of learned patterns in computations for higher-order language computations. One implication for L2 learning is that NLNC could interfere with the computations that are not compatible with the established patterns of the statistical or prosodic regularities.

Studies in second language acquisition have converged to support the NLNC hypothesis, showing that there is only a limited time window during which the neuronal circuits are more responsive to experience and make appropriate adaptation to form a typical way of processing. Beyond this critical period (Lenneberg, 1967) or sensitive period (Knudsen, 2004), the 'plasticity' dissipates while the neuronal circuits mature, therefore in the case of language acquisition, native-like L2 attainment is almost impossible (Birdsong & Molis, 2001; Toppelberg, 1997). With a promising theory (NLNC) that implicates a critical period and that predicts ultimate attainment in language acquisition, AoA seems to be a better determining factor than the learning context. The difference in the influence of AoA and learning context on second language learning was shown by Silverberg and Samuel (2004): semantic priming was found in early bilinguals, but not late bilinguals irrespective of their L2 proficiency. AoA is also thought to affect the shaping of the organisation of the lexicon network. There is evidence that words that are learned earlier in life have a processing advantage over words that are learned later in life and the age at which a word is acquired will influence the connections between that word and its corresponding meaning (Carroll & White, 1973; Gilhooly & Logie, 1982). One implication is that for late L2 learning, the form and meaning of L2 words will not be connected as strongly as those of L1 words.

Yet another way of categorising bilingualism is a distinction between *natural* acquisition and formal language learning (Schreuder & Weltens, 1993). Like the compound-coordinate dichotomy, it is not without criticism either. During bilingual acquisition, there are usually other sources of language input in addition to formal education, e.g., Dutch children can naturally learn English and German from the media, while formally learning them in school. The setting where languages are acquired presents rather different learning contexts, depending on what the language of instruction is, e.g., a second language can be taught through a first language or through an immersion programme (Swain, 1998). Central to the theme of this thesis is a potential relation between the nature of acquisition and the organisation of the mental lexicon or its effect on the degree of bilingualism.

It is not yet clear how AoA, learning contexts and the nature of language acquisition interact in shaping the bilingual lexicon organisation and in predicting the ultimate attainment of a second language or the degree of bilingualism (Kroll & Tokowicz, 2005). However, the finding that AoA leads to "a qualitative difference in how language is processed by the brain above and beyond language proficiency" (Kroll & Tokowicz, 2005), and that AoA accounts for fifty percent of variance in L2 ultimate attainment (Birdsong & Paik, 2008) encourage the use of AoA as a tentative protocol in profiling bilinguals as early vs. late bilinguals. Although this distinction, according to the NLNC hypothesis, postulates a biological difference between two types of bilinguals, it does not imply that late bilinguals can never achieve high L2 proficiency. The question is not whether late bilinguals could achieve native-like proficiency, but whether there is a qualitative (and quantitative) difference in L2 performance between a native-like bilingual and a native monolingual beyond age effect and language proficiency.

Second language speakers who are rated by a conversation partner as native-like have shown traces of non-native features when their speech is subject to analysis using linguistic instruments (Hyltenstam & Abrahamsson, 2000). Hyltenstam (1992) reports non-native grammatical and lexical features in bilinguals who learned their second language early in their life.

In a grammaticality judgement study, Schachter (1989) discovered that native Korean speakers who were learning English failed to recognise subjacency violations in English. J. S. Johnson and Newport (1991) tested Chinese-English bilinguals who were divided into three groups according to their AoA and showed a continuous decline in performance on subjacency recognition. Coppieters (1987) found both quantitative and qualitative differences between natives and near-natives, even though the two groups seem to be equivalent in proficiency. With an aim to replicate his study, Birdsong (1992) tested selected bilinguals who passed the native-like test and concluded that near-natives can be similar to natives. Similar observations were made by Marinova-Todd (2003) who identified 3 among 30 screened late L2 learners, who performed just like native controls across nine tasks.

Birdsong (2006) commented that language learners may be able to become indistinguishable from natives in performing some tasks, but not all. Areas like lexical retrieval, structural ambiguity resolution, and detection of acoustic distinctions in the areas of syllable stress, consonant voicing, and vowel length were where performances of natives and near-native learners deviate. Paradis (2009) associated slower and non-native-like performance of L2 speakers with their language *learning* as opposed to *acquisition*. The finding that bilinguals with all-round nativelikeness are so rare, was thought to be a result of rare cases in which L2 was acquired like L1 was. The not-so-native-like aspects of bilingual performance may be due to dispersed attention allocated to subcomponents of language processes that are not fully automatised or integrated into bilinguals' implicit linguistic competence.

To sum up, firstly, the chances that bilinguals attain native-like proficiency depends to a large degree on their AoA, among other things. Secondly, evidence has shown that bilinguals do differ in their L2 performance from that of the natives in both qualitative and quantitative terms. Even if the difference falls outside the perceivable range of a human rater, it has been picked up with analysis using instruments;

and even when near-native bilinguals achieve native standard of performance in quantitative terms, e.g., response time and accuracy, evidence shows that they differ qualitatively (Paradis, 2009).

A necessary condition in conceptualising how language interpreting is carried out by testing untrained bilinguals is that one must know how comparable the bilinguals in many bilingualism studies are with potential bilingual participants in the present study and be prepared to justify the choice of such participants. Table 2.1 lists a selection of bilingualism and interpreting studies found in the database of ISI Web of Knowledge [9] (search keyword: language translation) between the year 2005 and 2009. The purpose of this small-scale survey was to highlight: 1) participants' bilingual acquisition history, particularly their AoA, was not always available; 2) having acknowledged that studies differ in their research questions and participant qualification, most bilingualism or interpreting studies target highly proficient bilinguals as their ideal participants. This survey suggested that participants in many studies started learning their L2 at around the age of ten, which can be considered late bilinguals in terms of either cut-off point, e.g., 6 year-of-age, (Hyltenstam & Abrahamsson, 2003; Long, 1990; Unsworth, Heitz, Schrock, & Engle, 2005), or in terms of critical period, e.g., age between 3-8 (Meisel, 2008). Some researchers even explicitly labelled their participants as late or unbalanced bilinguals in bilingualism studies (Duyck, Vanderelst, Desmet, & Hartsuiker, 2008; Van Assche, Duyck, Hartsuiker, & Diependaele, 2009).

To a limited degree, participants who took part in some bilingualism studies can be quite similar to those who participated in studies that required overt translation (e.g., Dillinger, 1994; Macizo & Bajo, 2006). Given that the understanding of the organisation and operation of bilinguals' languages has been shaped by modelling the data collected largely from late bilinguals, it is likely that bilinguals and trained

[9] http://www.isiwebofknowledge.com/

translators only differ in where they fall on a distribution of a multivariate regression model that uses bilinguals' basic language skills that were outlined in section 2.3.3 as parameters into the equation of performance modelling. Christoffels, De Groot, and Waldorp (2003) modelled their Dutch-English untrained translators' working memory measures (English reading span, Dutch reading span, digit span), picture naming, word translation (forward and backward translation) against their simultaneous interpreting performance. They showed that picture naming and word translation were significantly correlated with the participants' interpreting performance, rated by independent judges on Likert scales which resulted in two scores. Their graphical analysis modelling suggested that "word translation efficiency and working memory formed independent subskills of untrained bilinguals". In summary, by asking just how bilingual they are, this small survey showed that bilinguals that took part in some bilingualism studies were comparable with those who took part in some interpreting studies, at least in their age of L2 acquisition.

Table 2.1: Bilingualism status of participants in recent bilingualism studies. N: sample size, NA: Not Available

Study	L1	L2	n	L2 AoA	Remarks
Gollan, Montoya, Fennema-Notestine, and Morris (2005)	English	Spanish	31	Early age	Undergraduates at the University of California at San Diego
Finkbeiner, Almeida, Janssen, and Caramazza (2006)	Chinese	English	4	NA	Self-rated proficient
	French		3		
	German		2		
	Italian		1		
	Japanese		1		
	Spanish		5		
Phillips, Klein, Mercier, and deBoysson (2006)	English	French	15	6.9	Self-rated highly proficient in L2
Ferré, Sanchez-Casas, and Guasch (2006)	Spanish	Catalan	49	1.1	Compound bilinguals
Schwartz and Kroll (2006)	Spanish	English	23	7	Ss rated themselves higher in L2 proficiency
Friesen and Jared (2007)	English	French	50	NA	Half were more skilled bilinguals
	French	English	50		
Prior and MacWhinney (2007)	English	Spanish	40	NA	Half were English-dominant
	Spanish	English	40		
Tokowicz and Kroll (2007)	English	Spanish	25	12	AoA is the average across Ss in 2 experiments
	Spanish	English	20		
Voga and Grainger (2007)	Greek	French	24	NA	Most Ss were from the Language Department
Basnight-Brown, Chen, Hua, Kostić, and Feldman (2007)	Serbian	English	44	10.1 12.1	Sample size is the result of proficiency screening to match proficiency between two groups
	Chinese		44		
Schoonbaert, Hartsuiker, and Pickering (2007)	Dutch	English	32	NA	They had formal instruction in English for at least 5 years
Thierry and Wu (2007)	Chinese	English	30	12	IELTS = 6 or 6.5
Christoffels, Firk, and Schiller (2007)	German	Dutch English	24	NA	Dutch and English proficiency was comparable

–Continued on next page

– Continued

Study	L1	L2	n	L2 AoA	Remarks
Grabner, Brunner, Leeb, Neuper, and Pfurtscheller (2007)	German	English	na	na	Ss were students of translation and interpreting at the University of Graz (Austria)
Basnight-Brown and Altarriba (2007)	Spanish	English	26	5	Writing, spoken comprehension, and speaking skills were self-rated as better in English
Duyck et al. (2008)	Dutch	English	18	NA	Authors categorised their Ss as unbalanced bilinguals
Hoshino and Kroll (2008)	Spanish	English	42	10.3 10.2	Spanish-English Ss rated perceived themselves more proficient in English
	Japanese		27		
Perea, Dunabeitia, and Carreiras (2008)	Basque	Spanish	40	NA	Ss acquired both languages since early childhood
Sumiya and Healy (2008)	English	Japanese	24	NA	Their length of studying or using Japanese ranged from 2 to 37 years
Ruiz et al. (2008)	Spanish	English	16	NA	All Ss were professional interpreters with 2 or more years of experience
Van Hell and De Groot (2008)	Dutch	English	16	10	
Van Assche et al. (2009)	Dutch	English	45	14.5	Ss were categorised as unbalanced bilinguals

To this point, it has been established that AoA is a useful criterion in profiling bilinguals. But being early vs. late bilinguals *per se* is not informative regarding what is to be expected from a bilingual in a translation task. Clearly, what needs to be in place is a more specific profile of bilingual participants and this is the objective of the second part of the two-stage analysis.

Decades of studies in second language acquisition, bilingualism, and language production offer a clear trend in approaching cross-language communication. It has been repeatedly shown why the mental lexicon plays a central role in understanding language processing including perception and production (Schreuder & Weltens, 1993), as it provides a bridge between form and meaning. "If one is able to gain a good understanding of how lexical access takes place, a major mechanism of speech production can be explained. Based on all this, it is no wonder that lexical encoding and the bilingual lexicon are the most widely researched areas of L2 speech production" (Kormos, 2006, p. 55). Characterising bilinguals' competence by considering the possible configuration of bilingual memory, e.g.,the size of lexicons and the links between each component, would then help us develop hypotheses of bilingual comprehension and production in a translation task. Parallel to the importance of the mental lexicon is an account for the execution of the grammatical rules. In relation to language comprehension, it needs to be clear how accessed lexical items are integrated to construct a mental representation. As regards language production, one needs to be clear how selected lexical items are put together in a correct order to form a grammatical sentence. This approach can explain how bilinguals coordinate their declarative and procedural knowledge in a task that requires language processing in a less familiar language. According to Ullman (2001), declarative memory system is implicated in the learning and use of knowledge about facts and events, whereas procedural memory system may be involved in the learning and control of motor or cognitive skills. Regarding language learning and use, it is posited that declarative memory underlies the learning and use of lexical knowledge whereas procedural memory system subserves the learning and use of grammatical knowledge. There is evidence from animal studies suggesting that procedural memory may be subject to critical period (Fredriksson, 2000), whereas in human studies, declarative memory could actually improve with age (Di Giulio, Seidenberg, O'Leary, & Raz, 1994). In declarative/procedural memory theory, age of exposure and practice play major roles on the degree L2 learners rely on procedural memory for grammatical computations (Ullman, 2001). The grammatical computations in late L2 learners were found to be more dependent on declarative memory, which was associated with worse behavioral performance at grammatical tasks. Although the evidence comes primarily from comprehension studies (e.g., Weber-Fox & Neville, 2001), one recent study found that late bilinguals differed from proficient bilinguals in the use of L2 morphology in subject-verb agreement production (Foote, 2010). Two candidate theories were available so far. One is that bilinguals' errors of subject-verb agreement in production may be a function of the relative morphological richness of languages. The other concerns the increased cognitive load associated with bilinguals' production in a less proficient language might lead to higher rate of errors or non-native-like performance. It appears that cognitive resource interacts with proficiency in influencing L2 speech planning, thereby less proficient bilinguals are more error-prone in an on-line task such as fragment completion. A full elaboration on how cognitive resource (e.g., working memory) might play a role in language interpreting is presented in section 2.4.4.

2.3.3 Bilinguals' linguistic profile: Lexicon, Access, Efficiency

The mental lexicon plays a central role in language processing because it bridges the form and meaning of lexical items. Figure 2.19 is a simplified sketch of the mental lexicon that illustrates two assumptions, which are also two recurring themes in this section. First, language comprehension and production share one mental lexicon, and second, lexical knowledge is multi-dimensional.

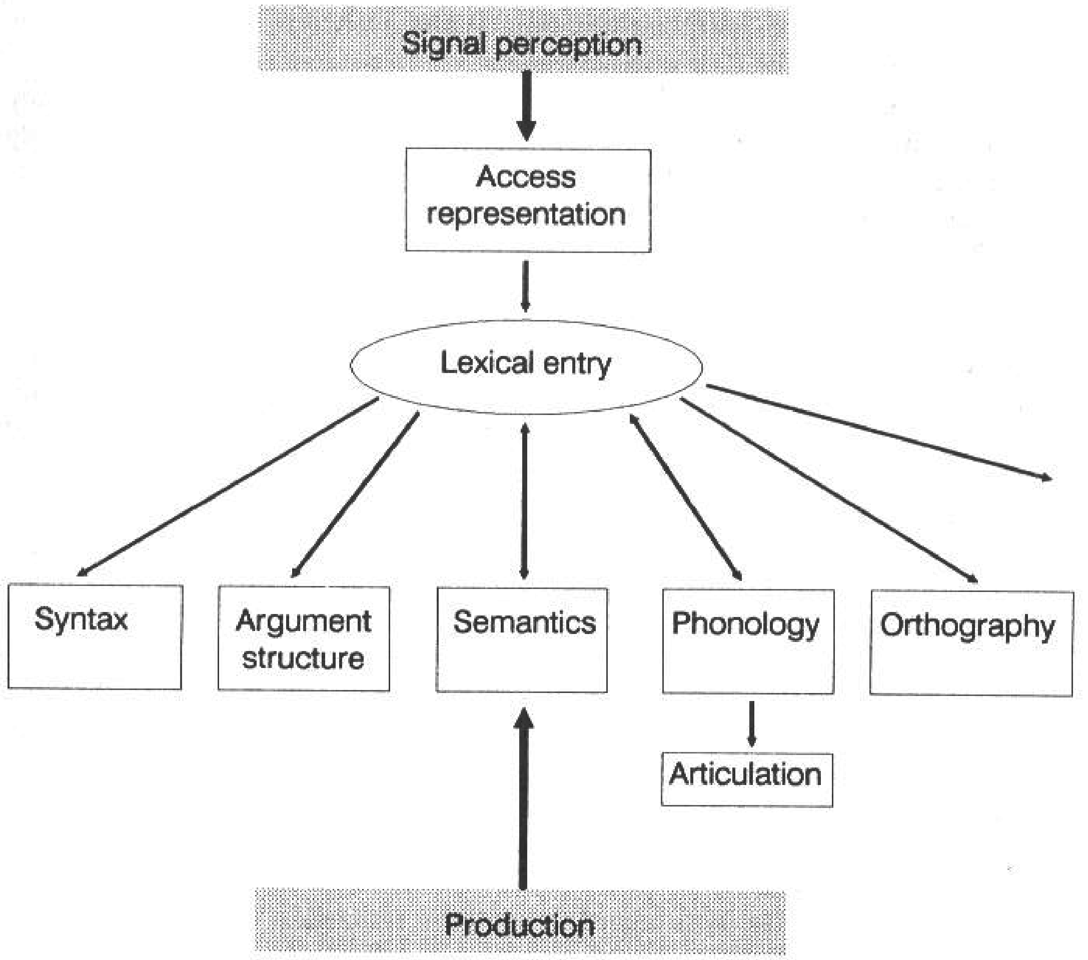


Figure 2.19: A representation of the mental lexicon. Adapted from Schreuder & Weltens (1993)

The mental lexicon's role in monolingual communication has been rigorously studied and recently became a hot topic in studies of bilingualism and beyond. One particular example was an attempt of direct application of a bilingual memory model to the research on simultaneous interpreting in a conference presentation (Kiyoshi, 2008). The model that appeared in the presentation was the Revised Hierarchical Model of Bilingualism (RHM). A potential problem in the direct application of this model is hard to miss: RHM was developed in modelling the configuration of bilingual memory at the lexical level, e.g., the extent to which two languages are separated in their lexicons and unified in their conceptions, but not at the sentence level. When dealing with a subject that involves sentence or even passage translation, a theoretically more viable approach would be placing a bilingual memory model in the framework of language processing, e.g., models of language comprehension or production, and discussing how processing at the lexical level interacts with or constrains subsequent processes that involve higher levels of processing. In this section, a bilingual memory model is brought to bear on the components of comprehension and encoding that were introduced in section 2.2.1. By following Kroll and Dijkstra's (2005) assumption that bilinguals use their bilingual mental lexicons in both language comprehension and production, this section will start by briefly reviewing the evolving models of bilingual memory. One model will be selected for later discussion, which is organised in terms of the directions of translation, i.e., L1 to L2 (forward translation) and L2 to L1 (backward translation). Discussion in each direction of translation will be centred around the selected bilingual memory model with an aim to characterise bilingual translators' competence in four aspects: the size of the bilingual lexicons, connections between the components of the bilingual lexicons, the automaticity in accessing bilingual lexicons, and the adequacy of activated information for communicative purposes in translation.

A straightforward question that can be asked is how a translator successfully converts the form in one language into concept and then maps the concept to the form in a different language. Imagine a scenario in which a translator (called 'translator T' in the rest of the thesis) is translating for a Chinese client but suddenly the translation is disrupted by a long silence. This translator's crisis is probably not a difficulty in conceptualising the Chinese word's meaning or deciding which language to select for production, but the unavailability of a lexical entry 保證 (guarantee). This unavailability of lexical item in a second language can also pose a challenge to translators when they translate from L2 into L1. Because a bilingual is thought to know more L1 words than L2, consequently each L2 word can usually be directly mapped onto its L1 equivalent, given that these L2 words were learned by associating each new L2 word with one or more L1 translation equivalent(s). But the reverse is not necessarily true due to smaller L2 lexicon relative to that of L1. However, translators who are highly proficient or early bilinguals may have a very different experience in translation than those who are less proficient or late bilinguals, because the former are more likely to have a balanced language competence. In order to account for the discrepancy between the size of two lexicons and the varying connection strength between L1 words and their L2 equivalents that were acquired at various points during a bilingual's language development, a satisfactory model of bilingualism has to feature developmental parameters in the size of lexicons and the network between lexicons and concepts. Among mainstream bilingual lexicon models, Revised Hierarchical Model of Bilingualism(RHM) (Fig. 2.20) is the most influential one that has both features.

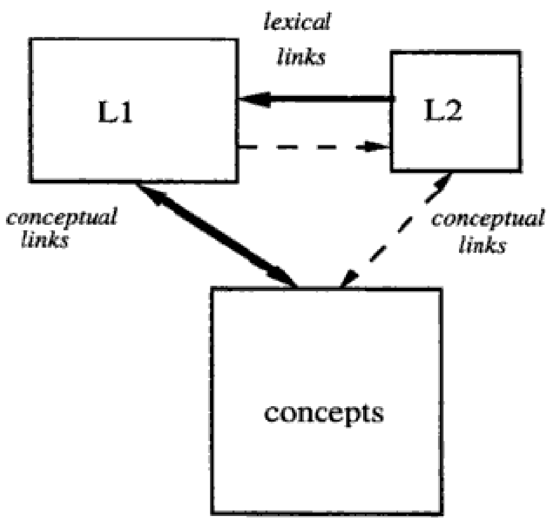


Figure 2.20: Revised Hierarchical Model of Bilingualism. Adopted from Kroll and Stewart (1994)

In Figure 2.20, two salient features of the RHM model are the asymmetry in the size between L1 and L2 lexicons and that in the connection strength and directionality between each component. The larger L1 box relative to L2 was an analogy illustrating that bilinguals, even when they are highly proficient, know more words in their L1 than L2 (Kroll & Stewart, 1994). In other words, every L2 word usually can map onto an L1 equivalent, but not vice versa when a bilingual's L2 lexicon is smaller than that of L1. Among the connections, L2-to-L1 lexical association (unidirectional dark line) was assumed to be stronger than the other way round (dotted line), because L2 words were usually learned by item-to-item association, mapping newly encountered L2 words onto the existing conceptual structure that was originally constructed with a bilingual's first language. The bi-directional connection between each lexicon and the concepts implies that bilinguals' lexicons share a unified concept system, but again the connection strength is stronger between L1 and the concepts than that between L2 and the concepts. These asymmetries in connection strength allow one to hypothesise that translating L1 into L2 is conceptually mediated, therefore it would be slower. In contrast, L2 words are more likely to be translated via the route of direct lexical association with their L1 equivalents, hence the process would be faster. It then can be predicted that Spanish-English bilinguals whose L1 is Spanish should find it easier to translate *house* into *casa*. But translation from *casa* into *house* may engage conceptual activation and take longer. For the same reason, L1-to-L2 translation is likely to be influenced by semantic context. Kroll and Stewart (1994) used the word translation paradigm and manipulated the semantic context by grouping experimental material in semantically related or unrelated lists. Among their Dutch-English bilingual participants, word translation latency was shorter in backward translation (L2-L1) irrespective of the semantic context. Critically, only forward translation (L1-L2) showed sensitivity to the context manipulation: translation was slower when words were in categorised lists, but not when they were in the mixed-category list. Since Kroll and Stewart (1994) first proposed the RHM in modelling the asymmetrical bilingual memory, attempts to replicate their work generated mixed results in studies that manipulated the context availability. There was no difference between the two directions of translation when to-be-translated words were accompanied with pictures (La Heij, Kerling, & Velden, 1996). Bilinguals with lower proficiency were not sensitive to word concreteness (De Groot & Poot, 1997). L1 distractor words, against RHM's prediction, interfered with L2 translation (La Heij et al., 1990). Nevertheless, more and more studies appear to converge gradually on a possibility that the asymmetry in the relation between lexicons and concept, that is central to the RHM, may be more symmetrical than first thought, especially when bilinguals progress towards more balanced proficiency in the two languages (Kroll, Michael, Tokowicz, & Dufour, 2002). This trend is acknowledged and reflected upon in the Modified Hierarchical Model (MHM) by Pavlenko (2009), in which both strong and weak connections exist between each lexicon and the concept, suggesting that forward and backward translation can both be conceptually mediated, unlike RHM's hypothesis that only forward translation is conceptually mediated. In the MHM, the route through which words are translated appears to depend on the category to which L1 and L2 conceptual representations belong and also on the availability of textual context (Pavlenko, 2009). In other words, in so far as L1 and L2 words share conceptual representations, the bilingual model is more symmetrical than asymmetrical.

Figure 2.21 shows that the MHM retains RHM's strength of the developmental progression in the lexicon size and connections between the components of the model, but crucially it also includes the notion of shared conceptual representations across languages. MHM's inclusion of language specificity in lexical access was argued by Pavlenko (2009) to be directly relevant to the performance in bilinguals' language production, especially in their weaker (or less-dominant) languages. To a certain degree, MHM also has the strength of another important bilingualism model, Distributed Feature Model (DFM), which readily accommodates the effect of concreteness in word translation between languages: bilinguals translate concrete

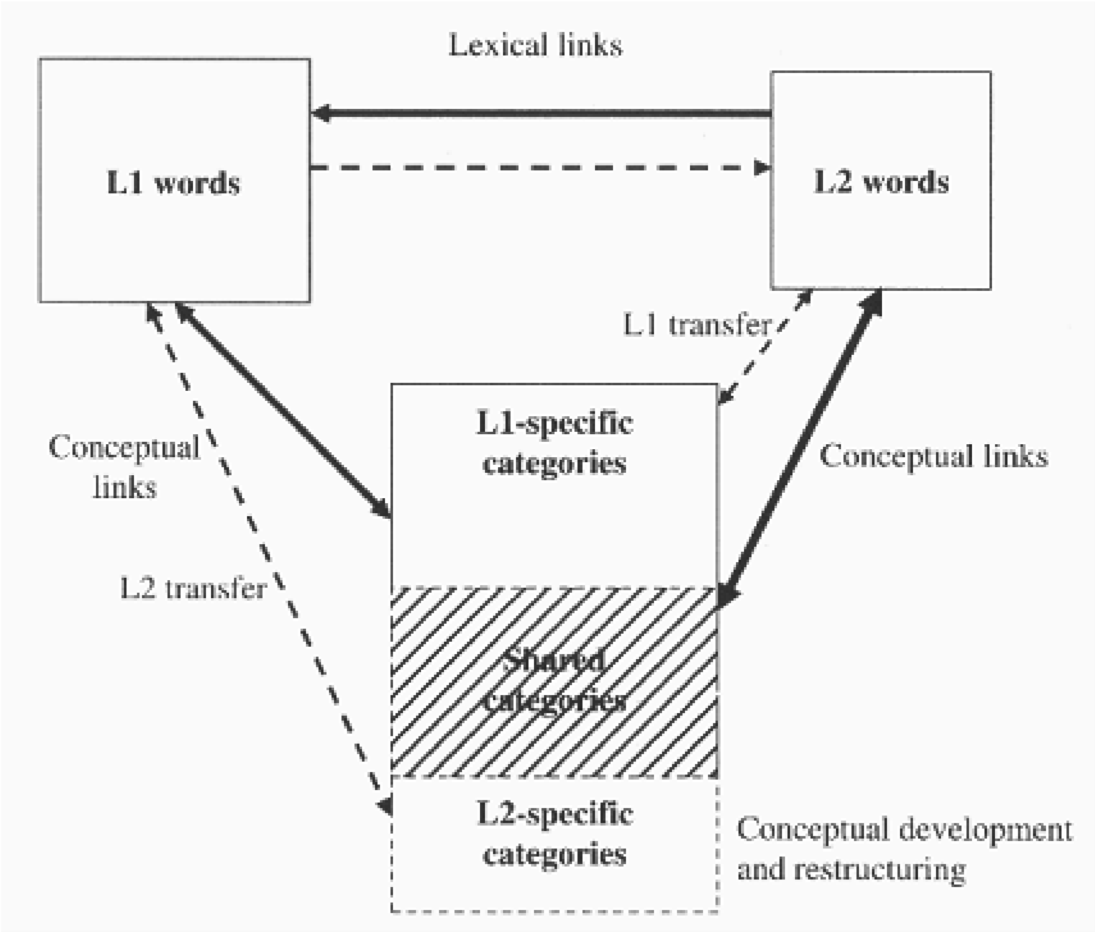


Figure 2.21: Modified Hierarchical Model, adopted from Pavlenko (2009, p. 147)

and cognate words faster than abstract words (Van Hell & De Groot, 2008; Kroll & Stewart, 1994), and this result was interpreted by the DFM as an evidence for more shared concept feature nodes between bilinguals' two languages for concrete words than abstract words.

Having said that, there are two reasons that led the present researcher to choose the RHM. Firstly the RHM has been empirically investigated for over a decade and its strengths are well-recognised in that it is parsimonious in formulating hypotheses and in interpreting experimental results. Secondly, the RHM can, in fact, incorporate the early vs. late bilingual distinction in at least two dimensions to profile bilinguals. Early bilinguals may have two lexicons relatively equal in size than late bilinguals. In addition, consistent with a growing number of studies that

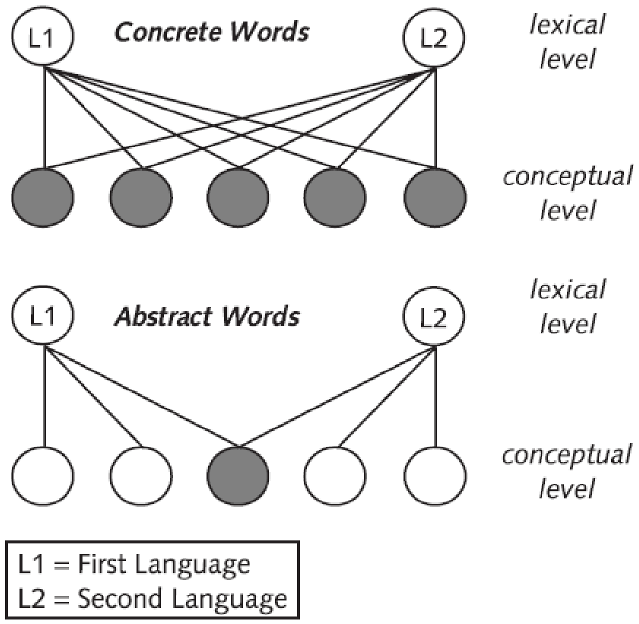


Figure 2.22: Distributed Feature Model, adapted from Kroll & Dussias (2004, p. 175)

report symmetrical rather than asymmetrical performances between forward and backward word translation, early bilinguals might have developed more symmetrical connections associating each lexicon and the concept than late bilinguals. The implication of this is that both forward and backward word translation can be conceptually mediated. And there is a tendency for a shift from lexical-mediated to concept-mediated word translation between bilinguals' languages when their bilingual proficiency becomes more balanced (Kroll, 1993).

But how can the RHM help to profile and relate bilinguals' competence to their performance in translation? Since the mental lexicon is the bridge between meanings and forms, and effective translation must be based on solid comprehension and at least comprehensible production, both are dependent on efficient use of a translator's mental lexicon. So the question can be addressed by considering bilingual memory as a store that supplies comprehension and production systems with data that are used as building blocks for constructing representations in comprehension or for constructing phrases in production. In line with the access view (Mezynski, 1983) that language comprehension is dependent on the accessibility of the vocabulary, and in turn, vocabulary access is dependent on fluency of lexical access and speed of word recognition, bilinguals' competence can be evaluated by examining their bilingual memory with regard to 1) how big the L2 *store* is relative to that of the L1 (which is essentially the asymmetry between two lexicons); 2) how robust the connections are between components of bilingual memory (which also informs how proficient and balanced a bilingual may be); 3) how automatic or fluent bilinguals access their bilingual memory; and finally 4) how useful the activated information is in meeting the demand of cross-language communication once a data point is selected, i.e., the completeness with which the selected or activated data is specified, e.g., grammatical information of a lemma.

Lexical breadth and strength

Lexicon size has been an under-developed area primarily because of the difficulty in defining word knowledge and the unavailability of reliable measures (Laufer, 1998). This under-development could be one of the reasons as to why the RHM does not specify any detail of its bilingual lexicons, except that one is bigger than the other. Laufer (1998) reminded us that vocabulary knowledge is not only about its size, but also about its depth. It is generally accepted that "knowledge of words may progress from superficial to deep... the learning of a word usually progresses from receptive to productive knowledge. Therefore, a word that can be correctly used should also be understood by the user, when heard, seen, or both. The opposite, however, is not necessarily true" (Laufer, 1998). When bilinguals' competence in terms of their vocabulary size of languages may be sketched by the RHM in the symmetry between two lexicons, bilinguals with similar L2 vocabulary sizes do not necessarily demonstrate similar performance in a task like composition, presumably because of the differences in other dimensions of word knowledge, besides their sizes.

Although the size (or breadth) of lexicon has been a key dimension in lexical competence, recently the multi-dimensional approach, represented by the *global trait modelling* (e.g., Henriksen, 1999) started recognising the importance of other dimensions such as quality (or depth) of lexical knowledge (Read, 2000), receptive-productive dimension (Henriksen, 1999), and automaticity of access (Laufer & Nation, 2001; Meara, 1996). This *global trait modelling* approach not only backs up the rationale of discussing bilingual lexicon in four respects outlined earlier, the importance of other dimensions apart from size has been highlighted by Meara (1996) by asserting that when L2 learners have developed a sufficient size of L2 vocabulary, the organisation of their lexical knowledge may become more important than its size. However, it has to be established first that bilinguals have built a sufficient size of vocabulary before any discussion on bilingual performance can be meaningful, so that any effect observed in cross-language tasks is attributable to bilingual participants' insufficient L2 word knowledge in addition to its size. This would at least require a threshold of L2 vocabulary size. As significant correlations between estimated vocabulary size and L2 comprehension in both reading and listening appear robust and consistent across studies (Henriksen, Albrechtsen, & Haastrup, 2004; Milton, Wade, & Hopkins, in press; Qian, 2002), proposed vocabulary thresholds show some variation. According to Nation (1990), 3000 word families (or 5000 words) is the minimal vocabulary size that should enable L2 speakers' comprehension of most subject-non-specific English texts. In a recent study, Nation (2006) suggested that 8000-9000 word families are required to "read a range of authentic texts." Staehr (2008) concluded that a vocabulary size of 2000 words is likely to result in above-average performance in reading, listening, and writing. According to the target set by the Ministry of Education of Taiwan, students should be able to have a vocabulary size of 5600 English words when they graduate from high schools (Yang, 2006). This figure was reported to provide an estimated coverage of 90% to 95% of most English texts (Laufer, 1992). Given that this target is achievable

Table 2.2: Strength of lexical knowledge measured by four tasks

Task Type	Active/Productive (Retrieval of form)	Passive/Receptive (Retrieval of meaning)
Recall	Turn into water _	When something melts, it turns into _
Recognition	Turn into water. a. elect b. blame c. melt d. threater	Melt. a. choose b. accuse c. make threats d. turn into water

and met, and these high-school graduates keep enlarging their vocabulary in higher education at the rate of 1600 word families each year (Laufer, 1998), they can potentially have a vocabulary size of nearly 10000 words, large enough to meet the threshold of the same figure proposed by Hazenburg and Hulstijn (1996), when they finish their first degree. Therefore, ideally, university graduates from Taiwan should be able to comprehend subject non-specific English texts without too much trouble. Although L2 lexicon size in Chinese-English bilinguals seems adequate, note that *knowing* words does not imply that bilinguals are able to use them at will, because bilinguals' lexicons are thought to be composed of active (productive) and passive (receptive) vocabulary (Melka, 1997). L2 speakers' vocabulary can be categorised in the strength of its meaning-form mapping for each lexical item, and the difference in the strength constitutes a scale of difficulty of four levels, which can be used to characterise the ease with which L2 speakers access the meaning or retrieve the form of a certain lexical item (Laufer, Elder, Hill, & Congdon, 2004). Laufer et al. (2004) referred to the ability to retrieve the word form as 'active' knowledge and the ability to retrieve the word meaning as 'passive' knowledge. Nation (2001) suggested that the active knowledge of L2 words can be further tested by cued recall, as in supplying an L2 word that best matches a provided word definition, and by a recognition task, as in selecting one word from options that corresponds to a word definition. The same tests can be applied to passive lexical knowledge. Table 2.2 shows examples in the four tasks representative of four levels of difficulty in form-meaning mapping in a L2 speaker.

By sampling 30 items from each frequency band (2000 most frequent, the third 1000, the fifth 1000, the tenth 1000 and the Academic Word List), two main results emerged from Laufer et al.'s (2004) Computer Adaptive Test of Size and Strength (CATSS). There was a main effect of frequency, such that low frequency words were more difficult to respond to than high-frequency words across recall and recognition conditions. Their results also supported the hypothesis of three, but not four, levels of difficulty in form-meaning mapping in the order from the most to the least difficult: active recall, passive recall, followed by two equally difficult tasks of active and passive recognition. When lexical size and strength are taken into consideration in the frame work of RHM, it is easier to recognise that it is not only the size of vocabulary that matters, but also the composition of a bilingual's lexicons – the proportion of passive and active vocabulary. Even though passive vocabulary is larger than active vocabulary in both L1 and L2 lexicons (Webb, 2008), it is conceivable that when passive vocabulary is much larger in proportion to the active vocabulary in L2 lexicon, the more likely it is that a bilingual or translator could encounter problems in L2 lexical retrieval, e.g., the unavailability of the word *guarantee* in the example in section 2.3.3.

In which example, this translator might fall into the category of late or unbalanced bilinguals in RHM's term where L2 lexicon is smaller than L1, and in his/her L2 lexicon, passive vocabulary is probably much larger than active vocabulary. However, his/her slow lexical retrieval may as well be a result of lower efficiency in his/her L2 relative to L1 skill. Lower efficiency in L2 processing can be understood in an approach different from the strength of lexical knowledge. Rather than asking whether or not the 'link' exists for form-meaning mapping, the dimension of automaticity in accessing lexical knowledge primarily concerns how efficiently lexical information can be accessed or retrieved, presupposing that the links between concept and lexical items have already been established.

Automaticity in lexical and grammatical information access

This section discusses an important topic, originally brought up at the end of section 2.2.2, by asking two questions. How automatic bilinguals are in accessing lexical and grammatical information, and whether or not this information is adequate for bilinguals to form representations in comprehension and to construct structures in production? As was emphasised earlier, lexical knowledge can be evaluated from several dimensions – breadth, depth, and automaticity. This multi-dimensional view is especially important since the competence in these dimensions could have a bearing on bilinguals' performance both in comprehension and production.

So far, the dimensions that have been put forward to characterise bilinguals' competence (AoA, bilingual lexicon symmetry, lexical knowledge) have potentially covered a wide range of parameters that can help to predict the trend of translation performance in bilinguals. But there is another factor that is not to be missed, which concerns the degree to which functioning in the translators' second language can be likened to other skills that show improvement after practice. This refers to the degree of *automaticity*. In fact, this factor has been hinted at earlier in discussing the RHM and MHM. The links between each component in these models are assumed to vary in their strength according to a bilingual's L2 proficiency. The trend of becoming less asymmetric in the speed of word translation between backward and forward translation when a bilingual becomes more proficient in L2 implies a possibility that the quantitative change of performance, e.g., efficiency in lexical access, could be a result of qualitative reconfiguration of the bilingual lexicon, e.g., connections between words and concepts become stronger. What the bilingualism models lack, however, is a potential mechanism that could account for the qualitative reconfiguration. In this section though, the focus of discussion is not the mechanism that underlies skill automisation (see Segalowitz, 2003, for a review). Rather, it is specifically discussed how automaticity in language processing has been conceptualised in L2 speech comprehension and production, and how bilinguals can be characterised by the degrees of automaticity as has been proposed by Kahnemann and Treisman (1984). Although the underlying mechanism of skill automatisation will not be discussed, the concept and definition of automaticity will be briefly introduced, because the misconceptions about automaticity are much prevalent in the studies of second language learning and bilingualism.

As mentioned, the reason that automaticity is central to bilingualism and language acquisition is that second language functioning has been conceptualised as a skill that follows a similar progression of acquisition, development, and procedurization of many other skills. An immediate benefit from an automatised skill, languages for example, is the compelling efficiency of functioning in L1 relative to that of L2. In the studies published between 1974 and 1993, DeKeyser collected fourteen criteria for automaticity (DeKeyser, 2001). Among these characteristics, fast, ballistic, effortless, and unconscious have been experimentally investigated with an aim to elucidate the concept of automaticity in the context of language use. In relation to 'fast' as one criterion of automaticity, Posner and Boies (1971) discovered that letter recognition required little attention and its processing was rapid and not subject to interference from other ongoing activities. However, speakers whose L1 differs from their L2 in their orthographies, e.g., Hebrew vs. English, showed slower response time in recognising English letters than their English L1 counterparts (Segalowitz & Hulstijn, 2005).

When automaticity is characterised by 'ballistic,' it means that the processing is unstoppable and involuntary (Neely, 1977). A well-known example is the automatic access to the words' meaning in a semantic priming paradigm. Favreau and Segalowitz (1983) extended Neely's LDT (Lexical Decision Task) paradigm to test the 'ballistic' hypothesis among bilinguals. In their primed lexical decision task, participants were shown lists of six pairs of prime and target words that were manipulated in their semantic relatedness and the stimulus onset asynchrony (SOA).

Their experiment was designed such that participants were deliberately trained to respond faster to unexpected and semantically unrelated words, e.g., respond faster to a target word TABLE after seeing a prime word FRUIT relative to a neutral prime. The reason that this effect was made possible is because in the training session participants were shown proportionally more semantically unrelated than related prime words in a list. When the SOA was long (1150 ms), a facilitation was observed when participants saw unexpected and semantic unrelated targets, but performance was slowed down when participants saw unexpected but semantically related targets, e.g., FRUIT as prime and APPLE as target. One of the crucial findings was that the priming effect in bilingual's L1 appeared to be modulated by the SOA: unexpected and semantically related target words facilitated lexical decision when the SOA was short (200 ms), but not when it was long (De Groot, 1984; Den Heyer, Briand, & Dannenberg, 1983; Neely, 1977). These contradictory results indicated that participants were not able to suppress the activation of semantically related concepts even when they were trained to predict semantically unrelated words. When longer processing time was available, controlled processing overtook automatic processing. But when processing time was limited, semantic information related to prime words was activated automatically. Another critical finding was the contrast between highly-proficient and less-proficient bilingual participants. Only highly-proficient, but not less-proficient bilinguals, showed the automatic, or 'ballistic' response in the within-language lexical decision task. Recently, Kotz (2009) replicated similar results using ERPs (Event-Related Potentials) combined with LDT, and supported the hypothesis that L2 proficiency modulates category semantic priming effect. Taken together, when bilinguals become proficient in L2, they access the semantic information of L2 stimuli automatically like native speakers do. Within-language priming demonstrates the 'ballistic' dimension of automaticity in lexical access, but a more relevant question is how this dimension relates to the theme of this thesis, translation. As regards language comprehension in translation, this question can be explored by considering not only how automatic words in source language are recognised when presented to a bilingual or a translator, but also the possibility of the co-activation of their translation equivalents in the translators' target language. The reason for this consideration is that translators always switch between languages, the consequence of the co-activation of target language words in the process of lexical access in source language recognition could be a facilitation effect for target language production when overt translation is required. In relation to language production in translation, the focus of the discussion is on the speed with which a lemma in the target language is selected. Both lines of discussion would require empirical evidence from studies of cross-language semantic priming presented in the following discussion.

By using positive priming (Frenck & Pynte, 1987; Grainger & Beauvillain, 1988), positive priming with an episodic recognition task (Jiang & Foster, 2001), negative priming (Fox, 1996), and forward masked priming (Perea et al., 2008), these studies established within- and between-language priming effect among bilinguals. Frenck and Pynte (1987) reasoned that if within-language semantic priming is a consequence of automatic access to the prime words' semantic information, then it would be theoretically significant if there is evidence suggesting that cross-language semantic priming effect is due to automatic semantic activation of shared conceptual representation of words and their translation equivalents. Frenck and Pynte (1987, p. 391) rightly pointed out that "this result [across-language priming] not only would give proof of the existence of a semantic link between two words of two different languages of the same nature as that existing between two words of the same language, but would as well invalidate all arguments that attribute across-language context effects to subject strategies, such as translation". Unexpectedly, Frenck and Pynte's (1987) data showed that cross-language priming was significant among less-proficient but not proficient bilinguals and it was only present in the long, but not the short SOA conditions (also see Grainger & Beauvillain, 1988, for similar findings). The results in Frenck and Pynte (1987) pointed to a possibility that the priming effect observed in studies using visible primes (SOA >200ms) might have resulted from strategic use of prime words when time permits, rather than automatic inter-lingual activation (Frenck & Pynte, 1987). One of the strategies is L2-L1 translation of prime words (Perea et al., 2008). The issue of strategy use was resolved to a certain degree when studies adapted masked priming technique, in which the short SOA (200 ms or shorter) was thought too short for subjects to notice the prime and for their strategy to take place (De Groot & Nas, 1991). By using masked priming technique, De Groot and Nas (1991), Perea et al. (2008), and Schoonbaert, Duyck, Brysbaert, and Hartsuiker (2009) established the cross-language semantic priming effect which was taken to support different bilingual memory models. De Groot and Nas (1991) attributed the cross-language priming effect observed in cognate word pairs to the automatic spreading activation from the lexical node of one word in one language to that of its equivalent in another language via shared conceptual nodes.

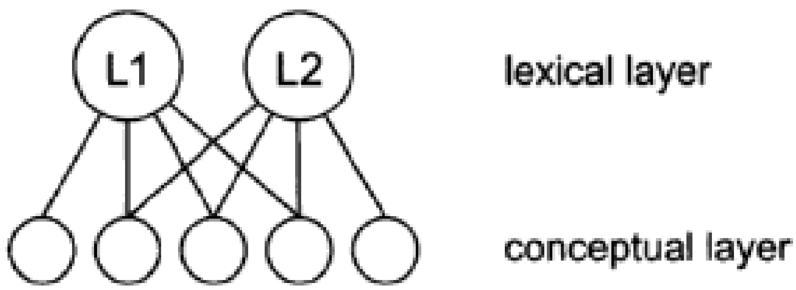


Figure 2.23: Modified Hierarchical Model, adopted from Pavlenko (2009, p. 147)

De Groot and Nas's (1991) null result in non-cognate word pairs was thought to be related to their participants' bilingual competence. Recent studies using masked priming technique, however, confirmed that the semantic priming effect can be observed in both L1-L2 and L2-L1 prime-target language combinations not only among highly proficient bilinguals (Perea et al., 2008), but also among unbalanced bilinguals (Schoonbaert et al., 2009). Schoonbaert et al.'s (2009) study is particularly significant regarding the role of bilinguals' proficiency, the cognate status of test materials and the nature of cross-language priming asymmetry approached from qualitative vs. quantitative perspectives. Schoonbaert et al. (2009) concluded that their results can be best accounted for by a refined model of DFM (Figure 2.24).

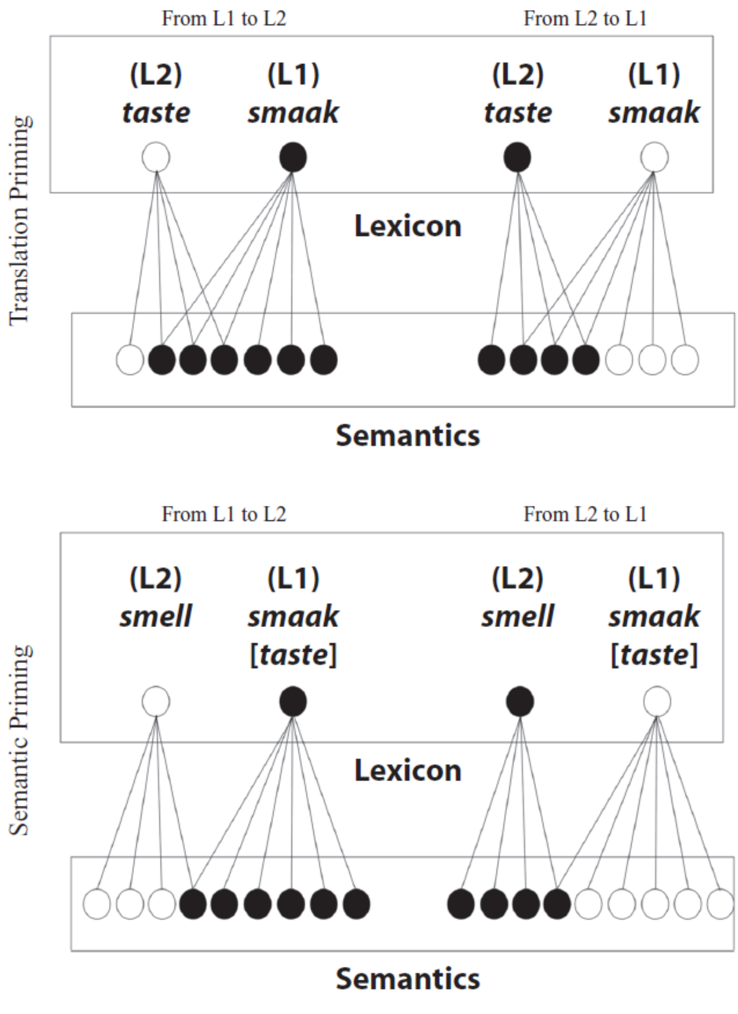


Figure 2.24: A refined model of Distributed Feature Model, adapted from Schoonbaert, Brysbaert and Hartsuiker (2009, p. 581).

In essence, the cross-language priming effect is hypothesised to be determined by the extent to which the conceptual feature is shared by two languages in terms of the proportion of shared nodes. Figure 2.24 shows the main assumptions made by the DFM: 1) bilinguals' conceptual representation is originally built with their L1, hence more nodes will be activated in the presence of a L1 words (e.g., smaak) than their translation equivalent (e.g., taste); 2) translation equivalent in a non-target language (e.g., taste) shares more conceptual features than a semantically related word (e.g., smell) does with a target-language word (e.g., smaak), as can be seen in the difference in the number of dark nodes between the two panels. Schoonbaert et al.'s (2009) results corroborated these hypotheses, but crucially the observation of a significant L2-L1 priming effect, although smaller in the magnitude relative to that in L1-L2, suggested that the asymmetry in cross-language priming appeared to be quantitative, rather than qualitative in nature. In other words, unbalanced Dutch-English bilinguals in Schoonbaert et al.'s (2009) study were able to access the semantic information of L2 words automatically. This conclusion may be true when bilinguals' languages share orthographic features, e.g., English and Dutch. But when they differ, e.g., English and Chinese, "the processing of words in a different script relies on other processes that are not as well practised as the processes of L1, so they take more time to complete" (Schoonbaert et al., 2009, p. 582). Their comment is compatible with work by other researchers that conceptual activation from L2 words can be weaker (Duyck & Brysbaert, 2004), and it builds up slower than the activation from L1 words (T. Dijkstra & Van Heuven, 2002). Slower word recognition in L2 could lead to general slow-down in sentence reading in L2, which has been attributed to "relatively less efficient use of low-level (e.g., phonological) information by the L2 reader" (Favreau & Segalowitz, 1983; Segalowitz, 1986; Segalowitz & Hebert, 1990). Segalowitz (1986) hypothesised two knock-on effects of less efficient processing in the lower level of language processing in L2 readers. First, processes that require attention will interfere with speech-based codes in memory for slower L2 readers. Second, less skilled readers might depend more than skilled bilingual readers would on the phonological codes in L2. Segalowitz and Hebert (1990) however discounted that the less-skilled L2 readers were more dependent on phonological codes than were fast L2 readers, and Segalowitz and Hebert (1990) suggested that slower readers could overload their working memory because their slower word recognition required working memory to hold the phonological representations of ambiguous sentential material until upcoming material may aid the semantic integration in sentence comprehension. Since resources are likely to be exhausted by holding the phonological representations of ambiguous material, non-phonological information could be lost. The loss of the non-phonological information could lead to poor performance in a task such as the judgement of semantic anomaly. Because translation utilises the meaning extracted from sentences by 'interpreting process' to accomplish other tasks, it might use 'post-interpretative processing' modulated by working memory (Caplan & Waters, 1999), just like the task of judgement of semantic anomaly would in Segalowitz and Hebert's (1990) study. The inefficient processing at lower level of sentence comprehension (e.g., ambiguity in word meaning) might also require working memory for semantic integration and result in the loss of content information.

Given that lexical access for word recognition can be automatic in the 'ballistic' sense in proficient bilinguals, would the same statement be true of lexical access for single word production in L2? In the reversed order of lexical access in word recognition, single word production starts from generating intended messages and continues by retrieving lexical items that correspond to the intended meaning. The Chinese-English translator in the earlier example would need to retrieve an item in English that matches his/her concept of the verb *guarantee* in a perverbal message. One heatedly debated question, coined as the *hard problem* (Finkbeiner, Gollan, & Caramazza, 2006), concerns how bilinguals select the right word in the right language since a word concept does not activate one but several candidates for production (Costa, 2004).

Most discussions of bilingual memory cannot avoid the subject of language specificity in lexical selection for single word production. It is generally agreed that this selection is language non-specific in nature, meaning that an intention of speaking a word in one language can also activate related words in the unintended language (Kroll, Sumutka, & Schwartz, 2005). Does this non-specificity apply to the translation task as well? It appears to depend on what type of translation is required, among other things. In the case of simultaneous interpreting, Grosjean (2008) contends that interpreters are in a bilingual mode, in which both languages are active. But in a task where source language input and target language output do not overlap in time, e.g., consecutive interpreting, translators perhaps are in the monolingual mode in one language during comprehension but in the other during production. Language mode is "the state of activation of the bilingual's languages and language processing mechanisms at a given point in time" (Grosjean, 2008). It can be best represented in two dimensions: the darkness of squares represents the level of language activation (the darker the square, the higher the activation) and the degree of bilingual mode is any point on the continuum ranging from monolingual to bilingual.

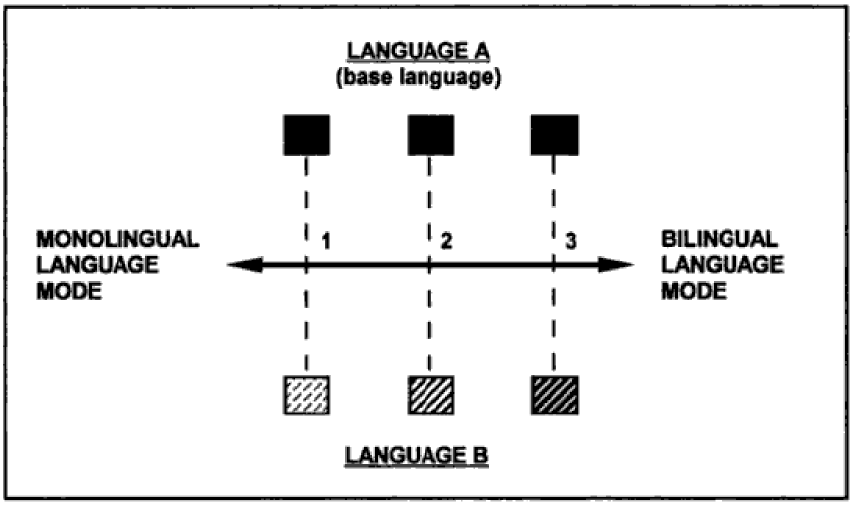


Figure 2.25: Visual representation of the language mode, adopted from Grosjean (2008, p. 70)

Bilinguals tend to shift their mode towards monolingual when their interlocutor has little or no knowledge of languages other than their own. On the contrary, bilinguals are more likely to be at the bilingual mode when they know that their interlocutor is also a bilingual and code-switching is acceptable (Grosjean, 1997). In a study designed to test Grosjean's hypothesis, participants' language modes were elicited by informing bilingual participants of the language capability of their interlocutors, and their task was to retell French stories that were littered with English words.

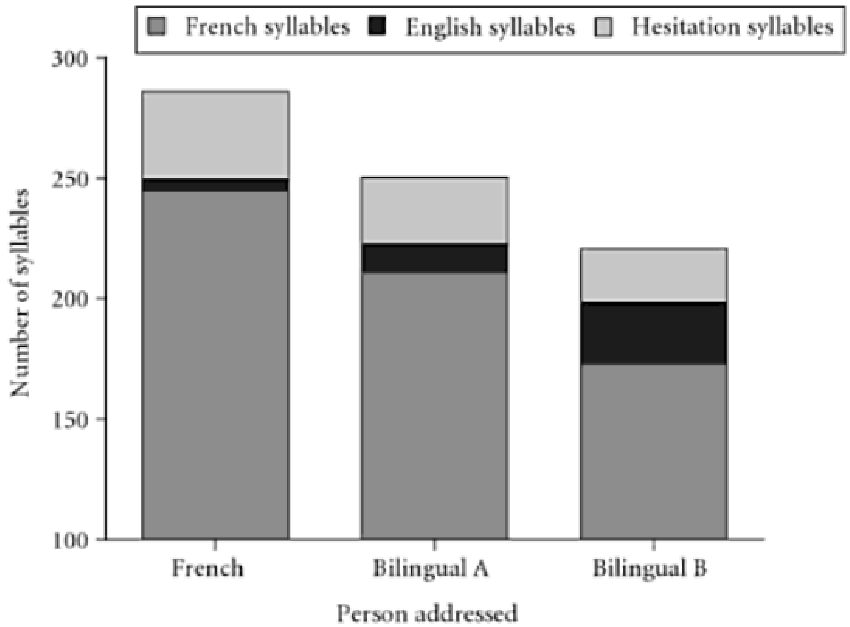


Figure 2.26: Production performance of French-English bilinguals as a function of language mode

Figure 2.26 shows that the mean number of French, English, and hesitation syllables in the story retelling task was a function of the type of person addressed. When the interlocutor was a French monolingual, participants used only a minimal amount of English words that probably were difficult to be translated into French in the testing material. The large amount of hesitations relative to other two conditions was interpreted as an effort in conveying the messages by translating those English words mixed in a story. If the effects of the interlocutor's bilingual status (monolingual or bilingual) and content (with or without code switches) apply to translation, a translator would set his/her language at a monolingual mode, presumably because his/her clients are monolinguals in different languages. But it would oversimplify the matter if the interlocutors' language and the message content were the only factors that affect the bilingual's language mode. Marian and Spivey (2003) suggested that an experiment itself can provide enough cues to move bilinguals away from their monolingual mode. "The fact that the bilinguals knew they were taking part in an experiment on bilingualism, that they were tested by bilingual experimenters fluent in both languages, and that the two languages were tested in adjacent experimental sessions," were likely to change the bilinguals' language mode (Marian & Spivey, 2003, p. 100) (Figure 2.27).

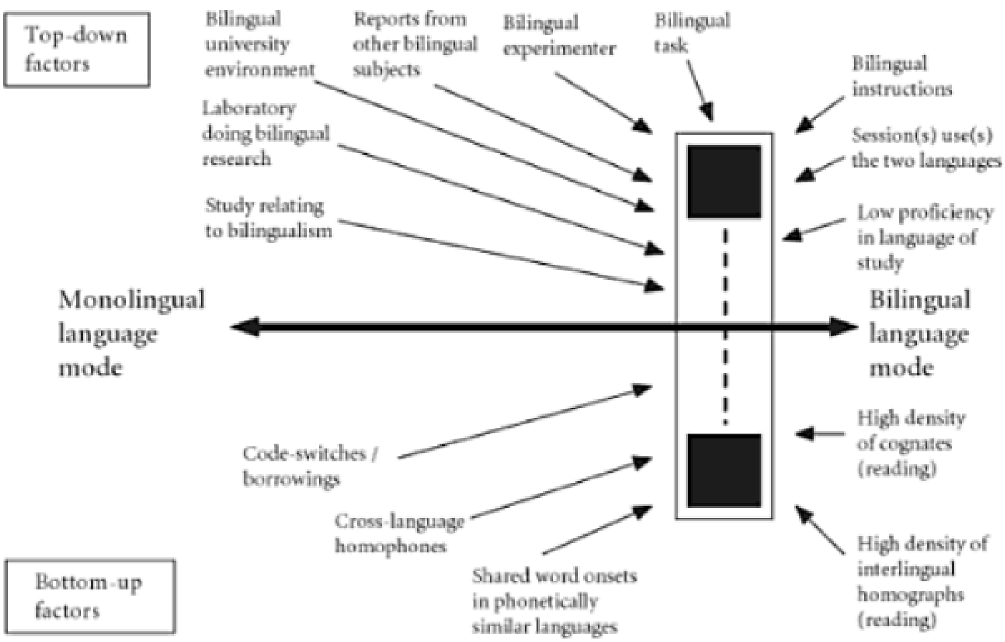


Figure 2.27: Factors that modulate language mode setting

With two factors promoting a monolingual mode and one that encourages a bilingual mode, these factors described above are apparently contradictory in biasing the bilinguals' language mode setting. How these factors reconcile is a relevant but not critical issue. Without complicating matters further, language selection is thought to be constrained by the task demand and context cues (Grosjean, 1997). Although it remains inconclusive as to at which stage in language production is the decision made on which language will be used, one of the popular views suggests that "the intention to speak in L1 or a second language is part of the preverbal message (LaHeij, 2005, but see an inhibition account and selection mechanism approach in Costa, 2004). Figure 2.28 illustrates that when a Dutch-English bilingual is required to name a picture of a dog, a concept needs to be selected and the preverbal message will specify to which language the sought-for lexical item belongs. This embedded language cue "ensures that the word in the intended language reaches a higher activation level than the translation equivalent in the non-intended language" (LaHeij, 2005). Apart from the sought-for word DOG that becomes activated, a spreading activation also activates DOG's translation equivalent HOND and semantically related word LABRADOR. But because of the language cue (+ Dutch), HOND is selected. In the case of language translation, the demand of alternating between two languages might require both languages be kept highly activated during the whole assignment, but a salient context cue is at the same time provided by an interlocutor's bilingual status, which often, but not always, is monolingual. This context cue could bias a translator towards the monolingual end on the continuum of language mode (Grosjean, 2001).

Given that the decision on which language to be selected is resolved, another decision still remains to be made. And that is how a translator selects one lexical item among co-activated potential candidates, e.g., select one lexical item among *dog*, *Labrador*, and *cat* in naming a picture of dog. This issue has to be analysed in at least two stages. First, the premise of the debate about lexical selection between candidates is that there exists a target lexical item that corresponds to a preverbal message. An overlooked aspect of bilingual production is that unbalanced bilinguals often have problems in lexical retrieval due to slow access or the fact that target items have never been acquired (De Bot, 1992). De Bot (1992, p. 434) commented "lexical retrieval problems are fairly rare in a unilingual non-aphasic speaker. For

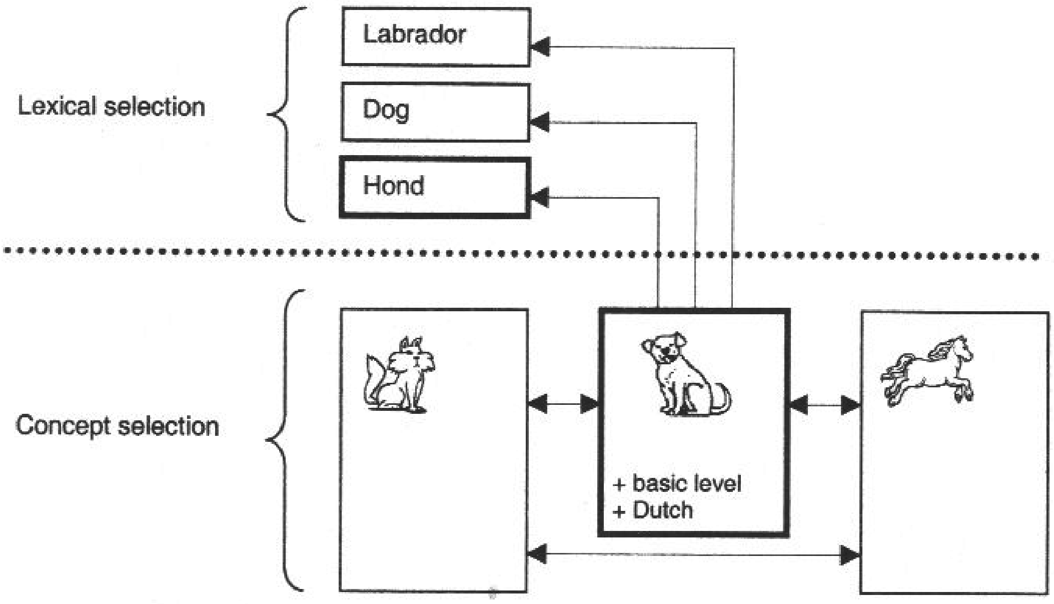


Figure 2.28: A bilingual lexical access model based on the complex access, easy selection approach in lexical selection of LaHeij (2005). Adapted from LaHeij (2005, p 303).

a bilingual speaker who does not have a perfect command of one of the languages these problems are commonplace, and the question is whether a bilingual speaking model can do without a mechanism that provides information about the availability of lexical items (i.e., both lemma and form characteristics) when the preverbal message is being generated." There is no easy answer to De Bot's question. It is possible that the conceptualiser does not know which lexical item is or is not contained in bilinguals' L2 lexicon (Poulisse, 1997). When an activation spreading from a conceptualiser to a translator's lexicons fails to activate any lemma in an intended language, or it only manages to partially activate a lemma, it is often necessary for a translator to regenerate a preverbal message or to resort to communication strategies, which can lead to a major delay in L2 speech production (De Bot, 1992). One of the known strategies is using generic terms, high frequency words that can approximate an intended lexical item, which is either not learnt at all or is on the translators' tip of tongue (TOT); or as D. Green (1993) supposed, a phrase of the same meaning must be computed. Although it remains unresolved as to what mechanism is responsible for monitoring and updating the availability of certain lexical items in the bilinguals' L2 lexicon, there is evidence suggesting that the bilinguals' vulnerability to lexical retrieval failure and higher rates of TOTs relative to monolinguals may be attributable to their lower frequency of L2 word use, which is associated with weaker form-meaning connections (Ecke, 2009), which in turn, is related to the bilinguals' competence that has been characterised in terms of their vocabulary size and the strength of lexical knowledge in previous paragraphs of this section. The process of lexical retrieval and the ease with which lexical items become selected can affect the production, most notably in its fluency, especially when the target language is bilinguals' or translators' L2.

When more than one L2 lexical item that matches a concept exists, the issue then becomes how a bilingual arrives at a final decision in lexical selection when there is more than one candidate item. For bilinguals with relatively larger and richer L2 lexicon, there may be a one-to-many meaning-to-form mapping relationship. So proficient bilinguals would have to select one item from a list of hyponyms (e.g., *Labrador, a Border Collie, German Shepard*, etc.) when less proficient bilinguals might just select a hypernym, e.g., *dog* in a picture naming task that presents them with a picture of Border Collie. Following from the debate of lexical selection in monolingual production, opinions are also divided on how L2 lexical selection works. Among bilingual production models, it appears that only two specify lexical selection in a L2 lexicon. De Bot and Schreuder's (1993) model[10] was an adaptation of Bierwisch and Schreuder's (1992) model (Figure 2.29, which was based on the production model of Levelt (1989). De Bot and Schreuder (1993) argued that a component VBA (verbaliser) between a conceptualiser and a formulator was necessary because "there is no one-to-one correspondence between concepts and words.

[10] Abbreviations in 2.29 CS: conceptual structure, VBL: verbaliser, INT: interpretation mapping, SF: semantic form, AS: argument structure, GF: grammatical form, PF: phonetic form, SS: syntactic structure.

The Conceptualiser does not present the formulator with a list of lexicalisable concept but rather with strings of information, consisting of conceptual primitives. The message fragments presented to the formulator often contain more conceptual primitives than can be expressed by means of one word, so they need to be cut up into lexicalisable chunks before lexical access can take place. Once the VBL has finished chunking, the right lemmas must be selected from the lexicon" (Poulisse, 1997). Lexical selection then follows the rule of *matching principle*, thereby "a particular lemma is selected only if it has all, and only those, of the primitives contained in the chunk-to-be-lexicalised" (De Bot & Schreuder, 1993).

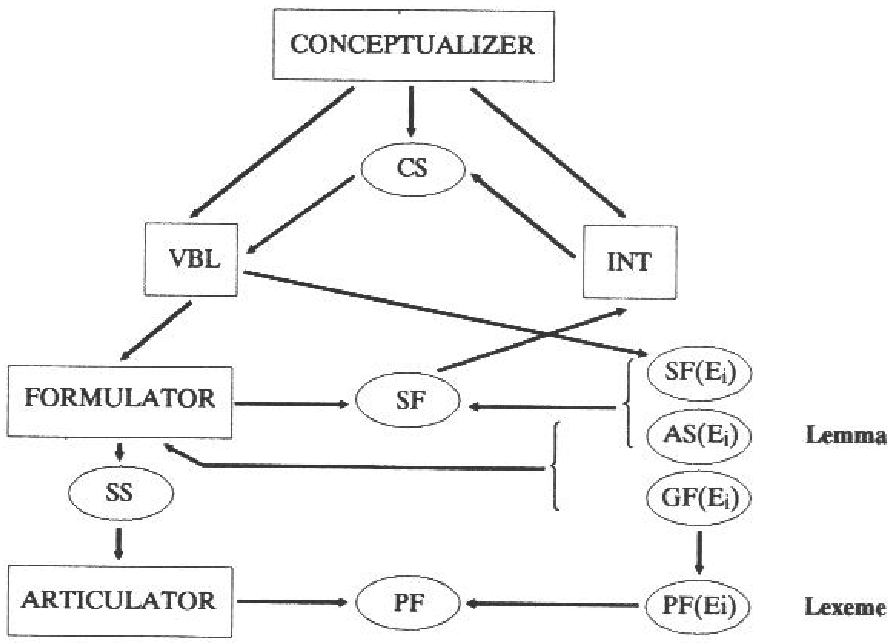


Figure 2.29: The interaction of differnt components in language production. Adapted from Bierwisch and Schreuder (1992).

But for the lack of empirical evidence and power of prediction of lexical selection, De Bot and Schreuder's (1993) model is limited to accounting for code-switch phenomenon. Nevertheless they highlighted one question that poses as a challenge to any bilingual production model: "how do speakers deal with the imperfect match between concepts and an L2 lexicon lacking fully specified lemma?" (De Bot & Schreuder, 1993, p. 212). This question will be discussed later.

Similar to the 'complex access, simple selection' principle in LaHeij's (2005) theory of lexical access, Poulisse and Bongaerts (1994) proposed that the preverbal message contains all the relevant information needed in lexical selection. The spreading activation from a conceptualiser will activate not only the intended lemma BOY, but also others that share semantic features with BOY. Figure 2.30 shows that each lemma receives different degrees of activation because of the preverbal message's specification in conceptual features and intended language. The result of this is that the lemma BOY receives more activation than other lemma within or across languages.

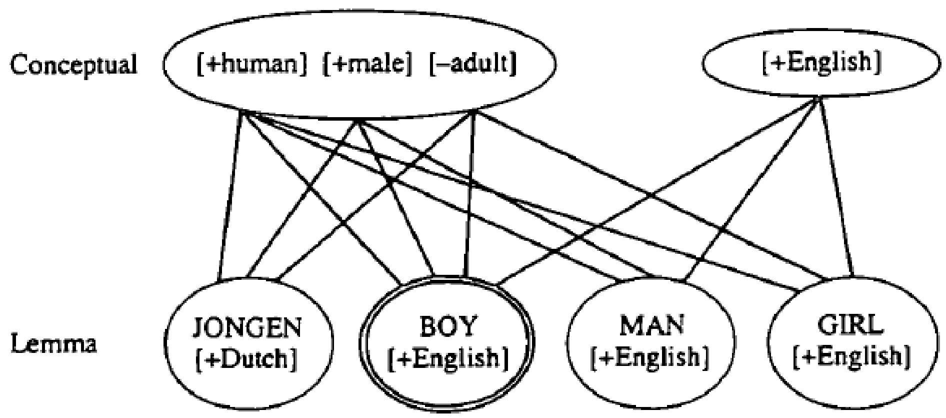


Figure 2.30: The selection of an L2 lemma through spreading activation. Adapted from Poulisse and Bongaerts (1994).

A major hypothesis formulated in their theory is that L2 speakers of different proficiency differ in the resting level of lexical activation. Less proficient L2 speakers tend to be much more dominant in their L1 and the occurrence of L1 words is more frequent than that of L2, therefore the resting activation level is generally higher for L1 than L2 lexical items. When language production is required in a task, an L1 lemma at a higher resting activation level can thus be selected more easily than its L2 equivalent, hence an instance of code-switch (i.e., intentional or un-intentional use of a non-target language). This hypothesis was corroborated with their data collected among three groups of Dutch learners of English (including university and high-school students). The number of switches was a function of word category (content words <function words) and language proficiency (university students <grade 11 <grade 9). Poulisse and Bongaerts (1994, p. 53) considered an account of reduced automaticity of speech production in the case of beginning learners, propounding "it could be that these learners had too little attention to spare for accessing and monitoring function words, which do, after all, convey little meaning." Their finding, however, still leaves the question about selection among target-language candidates unanswered, because code-switch is concerned with between-language choice, e.g., BOY vs. JONGEN (Dutch) in Figure 2.30.

It appears that existing bilingual production models have not been able to pin down how a concept links to a correct word, therefore no evidence is available to inform as to how automatic lexical selection is. Since most bilingual production models were based on Levelt's model, LaHeij (2005) suggested to take a step back to consider one major hypothesis in Levelt's original model, in which the retrieval of words is in a certain sense automatic. The major determinant of how accurate lexical access can appear to be the completeness of the specification in the preverbal message. When it contains all information or features required for a target word in a lexicon, the rest of the process is assumed to depend on the level of activation of each potential candidate lexical item: the one that receives the highest activation wins out. In this way, lexical access for production in L2 is similar to that in L1 and the most demanding and resource consuming step is conceptualisation. What would differentiate the efficiency with which words are accessed could be the bilinguals' L2 proficiency in terms of the strength of their lexical knowledge as mentioned earlier. To the extent that activation-based selection is valid, the difference in the bilinguals' strength of lexical knowledge, manifested in the domain of active vs. passive use of vocabulary, is likely to affect the strength of activation received by co-activated items within and across languages. Thus among less proficient bilinguals, "the process of using activated concepts to access L2 words may require the additional ability to resolve cross-language competition from alternatives in L1 with similar meanings" (Kroll & Tokowicz, 2001, p. 57). In this sense, lexical selection may be constrained by bilingual proficiency, and cannot be assumed fully automatic.

Recall that in an example in section 2.3.3 it took translator T a considerable amount of time to retrieve the verb *guarantee* in his L2 lexicon in a translation task. This slower lexical access in L2 relative to that in L1 is typical of unbalanced bilinguals, but lexical access is not the only dimension in characterising a bilingual's competence or psycholinguistic profile. Another dimension is exactly what is activated once a lemma is selected. By extending this consideration to a bilingual context, a question can be phrased 'does a correctly accessed lexical item promise an appropriate interpretation of a word and guarantee successful encoding in language production?'

Bilingual Memory: What is activated in lexical access?

Consistent with the assumption made by Kroll and Dijkstra (2005) that language comprehension and production rely on the same system, translators are assumed to use the same bilingual lexicon in decoding and encoding messages during translation. However, their bilingual memory may not serve all communicative purposes equally well. Bilinguals' lexical knowledge in breadth and depth differ in their languages, therefore, when their linguistic competence may be sufficient in the comprehension of L2 discourse, their relatively smaller productive vocabulary and weaker concept-meaning connections of a bilingual lexicon might hamper their performance in language production, as demonstrated in the example of translator T. On the contrary, translator T should not have much difficulty when translating into Chinese, not only because Chinese is the L1, but also because reading or listening comprehension does not necessarily involve lexical access, as contextual and other top-down processes might help solve comprehension problems (Kirsner, Lalor, &

Hird, 1993). The role of bilingual lexicon in translating between languages would, to some degree, depend on the task requirement and on whether the accessed lexical item supplies adequate information for language encoding. This line of reasoning is pertinent to the components in translation outlined in section 2.2.1 in 1) what is left highly activated in the translators' memory from parsing an SL sentence; and 2) how this memory might influence the way in which TL sentence is grammatically encoded. Note that comprehension in translation is only part of the entire task of translation, i.e., it requires prompt production to complete the whole task. So it is critical to study the switching between accessing translators' bilingual lexicon for comprehension in one language at one moment, and accessing the same system for production in another language the next moment in language interpreting.

An earlier discussion in the paragraph **Automaticity in lexical and grammatical information access** has reviewed studies focusing on the automaticity of mapping between form and meaning. In bilinguals, language-nonselective lexical access for both word recognition and production has been suggested to activate words in both intended and unintended languages. An across-language semantic priming effect (see paragraph **Automaticity in lexical and grammatical information access**) is usually taken as an evidence of bilinguals' access to the shared semantic information of the prime and target words. Logically, this priming effect cannot be interpreted such, that bilingual participants not only access the meaning of a prime word, but also access the translation equivalent of the prime word. Studies using several different paradigms, however, have provided support for language-nonselective access not only to the words' semantic representations but also their orthographic and phonological representations (see Dijkstra, 2005, for a review). A recent study that used both behavioural and physiological measures discovered that Chinese-English bilinguals automatically translated English words into Chinese in a strictly monolingual task context (Thierry & Wu, 2007).

Co-activation of translation equivalents in the non-target language when presented with words in a target language has been identified in studies using different production tasks. Fluent Dutch-English bilinguals were slower in naming pictures in English (L2) when distractor words, e.g., bench, were phonologically related to the target word's (e.g., mountain) translation equivalent, e.g., berg (Hermans, Bongaerts, De Bot, & Schereuder, 1998), suggesting that translation equivalents compete with target words for selection. Gollan and Acenas (2004) elicited TOT states among Spanish-English and English-Tagalog [11] bilinguals and found fewer TOTs in picture naming trials when the target words were cognate than when there were non-cognates. This result was taken as evidence for the phonological activation of translation equivalents. In a phoneme monitor task in which Catalan-Spanish bilinguals had to decide whether a certain phoneme was in the Catalan name of a picture. Pictures were manipulated so that a phoneme could be part of a picture's Catalan name, or part of its Spanish translation, or absent in either. It took longer for the participants to respond when a phoneme was part of the Spanish translation equivalent than in a control trial. Similar to the previous two studies, Colomé (2001) concluded that words in both languages were activated and it was the cross-language interfering that delayed participants' responses. This co-activation or competition account is not without criticism because of the opposite effect from translation equivalents of target words. One of the sources of objection or reservation came from the finding of facilitation effect of translation equivalents in a picture naming task (Costa & Caramazza, 1999). Costa and Caramazza (1999) suggested that "our results support the predictions derived from the language-specific hypothesis: the identity facilitation between languages indicates that the two languages of a bilingual do not compete during lexical access". In fact, language-specificity in lexical access is exactly the view taken in the earlier discussion on how bilinguals'

[11] Tagalog is spoken in the Philippines. It is related to Indonesian, Malay, Javanese and Paiwan of Taiwan (Wikipedia, 2009).

lexical access may be constrained by task context and the bilingual status of the interlocutors.

This view, however, cannot rule out the possibility of competition for selection between a target word and its translation equivalent in a translation task because translation differs from picture naming, most significantly in the source of intended message. To-be-named picture provides only an abstract concept (i.e., naming is conceptually driven), and picture naming does not involve language switch. In contrast, a source-language word provides translators with both a concept and a *name*, and the alternation between source and target languages in translation may incur a switch cost, the locus of which is generally thought to lie in the inhibition of a source language that might compete for selection in favour of production in a target language. In translation, to the extent that the recognition of a source-language word involves phonological recoding, similar to word naming in the same language (Frost, 1998; Seidenberg & McClelland, 1989), it follows that when word translation into a target language is required at the next stage, it naturally creates a language switch condition and might therefore incur a switch cost. Of great relevance to this thesis are the asymmetry in the switch cost and its implications in bilinguals' language control. The asymmetry of switch cost in picture naming that requires participants to alternate between languages suggests that more inhibition is needed to suppress L1 when production in L2 is required, but the consequence is that when L1 is required in a next trial, it takes longer to be re-activated. The opposite is true for switching to L2 because L2 is easier to suppress and re-activate than L1 (D. W. Green, 1998; Verhoef, Roelofs, & Chwilla, 2009). Note that when levels of proficiency and degrees of dominance often correlate, so that a bilingual does not need to be dominant in L2 to be able to produce high-proficient performance. According to Heredia and Brown (2004), bilinguals could be more dominant in L2 when L2 lexicon is used more often than L1. It is therefore equally plausible that a bilingual can be dominant on specific topics in L2, but remain otherwise less proficient in L2 on average.

A few interrelated questions can be asked. One concerns whether this switch cost is also present between L1-L2 and L2-L1 words translation. And the other is how bilinguals suppress the activation of one language while using the other. If indeed a mechanism must be in place to suppress one of the bilinguals' language during translation, what cognitive construct may be implicated? One candidate could be working memory (Michael, Dijkstra, & Kroll, 2002). Michael et al. (2002) suggested that "working memory is critical for controlling the activation of two languages, particularly in a task (like translation) that forces both languages to be active." Their result appeared to support the hypothesis, in that higher working memory span participants were more accurate than their lower-span counterparts in forward translation, but they did not differ in backward translation. It appears that bilingual functioning in both word recognition and production rely on the same bilingual lexicon, and both lexicons become active when bilinguals access a lexical item for recognition or production. This conclusion would correspond to an issue brought up in the paragraph **Automaticity in lexical and grammatical information access** about the possibility of coactivation of lexical items in bilinguals' non-target language. The rest of the discussion in this section will turn to the question, 'whether a correctly accessed lexical item guarantees an appropriate interpretation of a word and successful encoding in language production?'

Continuing the example of translator T, as the story has unfolded so far, his/her translation was disrupted with a long silence within a clause. This could be a sign of problems in accessing the target lexical item, the English equivalent of 保證 (guarantee). In order to use this example to illustrate the argument, here is another imaginary but reasonably possible scenario. When translator T at last retrieved the word, his/her translation went '*The stock broker [unfilled pause] guarantee at* *least ten percent of net profit growth to an investor.*' The ungrammaticality of this sentence may not cause a great difficulty in comprehension, but it can be a sign of various problems in this translator's grammar acquisition and language proficiency. Recall an earlier discussion of bilingual lexical access. To the extent that 'complex access and simple selection' holds in bilingual word production, a preverbal message with required specification should lead to the selection of a correct target word. The un-inflected verb *guarantee* suggests that the morphological feature for verb tense is possibly unspecified in the first stage of word production. The erroneous prepositional-object construction with the word *guarantee* shows that possibly, the syntactic information that is linked to the lemma GUARANTEE is underspecified, or it is a product of over-generalisation from other ditransitive verbs that can alternate between double-object and prepositional-object constructions, e.g., *give*, or indeed this particular syntactic knowledge has never been acquired in the first place.

It has been suggested that complete L2 lexical acquisition needs to include 1) phonological, 2) orthographic, 3) syntactic, 4) morphological, 5) semantic, 6) pragmatic/sociolinguistic, and 7) idiomatic information of each item (Nation, 1990, 2001). Kormos (2006) suggested that "until learners acquire these aspects of lexical knowledge, they might transfer this information from L1." By transfer, a straightforward definition is the "influence of L1 on L2 acquisition, language use, and comprehension (Kormos, 2006)" (see Odlin, 2003 for a review). One example is the transfer of diacritic parameters such as gender value between French and Italian, which are linguistically related languages. When the linguistic distance becomes larger, a great deal of diacritic information such as countability status, tense marker of verbs, plural markers of nouns, gender of pronouns can be entirely absent in a morphologically poor language, such as Chinese. This renders even transfer of inflection rules impossible from Chinese to English, and consequently increases the chances for less proficient bilinguals to produce a non-inflected verb in the example translation. So what lexical knowledges is required of bilinguals in order to produce grammatical sentences?

A language production model could potentially shed some light with regard to syntactic knowledge requirement. According to models of language production (Pickering & Branigan, 1998; Roelofs, 1992), nouns and verbs can be understood as nodes in a hierarchical network. At the highest level, conceptual stratum is where a message is generated. In the middle level, lemma stratum that contains lemma nodes is connected to the conceptual stratum and the word form stratum which contains morphology and phonology information. Each noun lemma node in Roelofs's (1992) model is connected to category (e.g., noun) or feature nodes (e.g., gender). The 'inherent activation' principle assumes that when a noun lemma is selected, its connected category and feature nodes become activated too (Pickering & Branigan, 1998). A similar idea was extended to verb lemma, but Pickering and Branigan (1998) posited that a root lemma (e.g., GIVE) is connected to various tense nodes, e.g., GAVE and GIVES. Figure 2.31[12] shows that each verb node is linked to feature nodes that specify aspect, tense, and number. Pickering and Branigan (1998) also assumed a combinatorial node, which specifies ways of construction available to the production system. For instance, a transitive verb *give* can be used to construct a double object (DO) or a prepositional object (PO) phrase. Using the verb *give* in a DO construction is believed to involve the activation of NP-NP node. It is also assumed that combinatorial nodes are shared between different verbs, e.g., *give* and *send*. So "prior activation of a combinatorial node (e.g., NP-PP), together with a lemma node (e.g., GIVE), should affect its likelihood of being activated in combination with any other lemma node linked to it" (Pickering & Branigan, 1998).

Although this model is most often used to illustrate the possible mechanism underlying structural priming, it is proposed here that it could also be useful in relating

[12]The labels T, A, and N refer to tense, aspect, and number.

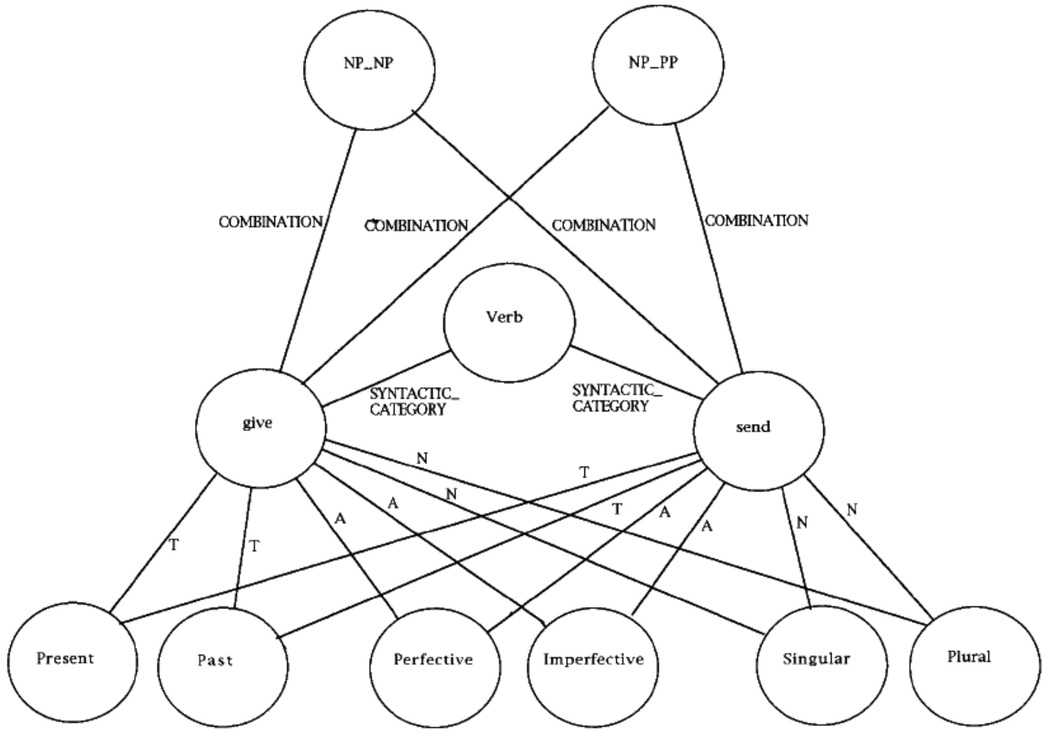


Figure 2.31: A partial model of a language production model. Adapted from Pickering and Branigan (1998).

bilinguals' lexical competence to their performance. Recently, McDonough's (2006) used syntactic priming to address the question of whether structural priming could be observed in a conversation paradigm between L2 speakers. There was a priming effect of PO, but not DO, in Experiment 1 that used both PO and DO prime sentences. In Experiment 2, only DO sentences were used as primes, but participants produced more PO constructions than DO. Intriguingly, the null effect of priming after DO primes is not limited to L2 speakers (Branigan, Pickering, & Cleland, 2000; Gries & Wulff, 2005). The extremely few instances of DO construction in McDonough's (2006) study led him to consider that participants' use of DO construction was item-specific, and this could stem from the reason that "these L2 participants might not have had complete knowledge of the morphological, semantic, and discourse constraints on dative alternation... in other words, they might have lacked the abstract syntactic information that could have been activated by the confederates' primes" (McDonough, 2006, p. 197). Although it is not clear what he meant by the 'abstract syntactic information,' some clue may be found in Figure 2.32.

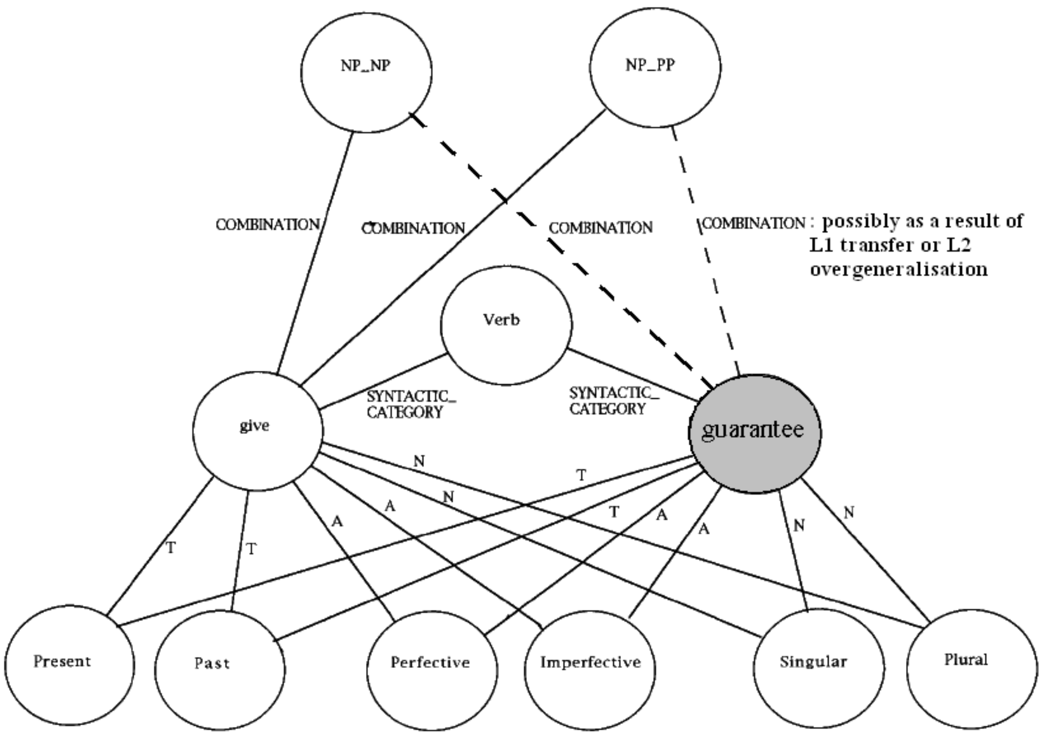


Figure 2.32: Incomplete acquisition of lexical knowledge might lead to unstable connections (dotted lines). Adapted from Pickering and Branigan (1998).

As shown in Figure 2.32, bilinguals' lack of full lexical knowledge, including rules that govern dative alternation (Mazurkewich & White, 1984), may lead to unstable or non-existing connections between a lexical item like *guarantee* and other items that share combinatorial nodes of NP-NP and NP-PP. Because lexical items like *guarantee* may not be linked to the combinatorial nodes, the highly activated node immediately after processing either construction is less likely to be reused in a construction using a target word, e.g., *guarantee*, hence a null effect in priming. In the case of translator T, it is possible that the lexical access to the lemma *guarantee* activated only partial feature nodes, but no nodes that specified the argument structure were activated or available, thus translator T had to transfer L1 grammar to L2 usage or had to overgeneralise the knowledge of familiar L2 items to the sentence construction using the verb *guarantee*. It has been shown that this overgeneralisation would be corrected when L2 speakers or L1 children perceive more correct use, most effectively through feedback from their interlocutors.

In summary, it has been argued that currently available evidence established that bilinguals' lexical information at different levels in a non-target language can become activated in the presentation of a target language lexical item. On the one hand, concomitant with the language-nonselectivity in bilingual functioning is a need for a regulating mechanism to control the selection and inhibition of languages. And on the other hand, selected lemma in lexical access for production does not always provide adequate information, such as options of argument structures, the consequence of which can often be ungrammatical phrase production.

In a two-stage analysis in characterising the translation competence of bilinguals, bilinguality was conceptualised in different terms, depending on the AoA (early vs. late bilingual), and the context of language acquisition (compound vs. coordinate bilingual) in the first stage of analysis. The AoA appeared to be a good predictor and criterion, for it has been supported by not only potential neural correlates in the framework of NLNC (p. 14), but also by studies in second language acquisition that were able to distinguish performance between early and late bilinguals. By comparing the performance of highly proficient bilinguals with that of native speakers, current evidence suggests that native-like L2 processes still differ from those of L1 in both quantitative and qualitative terms. In the second stage of analysis, this difference was further decomposed in bilinguals' linguistic profile, based on the analysis of their lexical knowledge. By looking into each dimension of bilinguals' lexical knowledge, section 2.3.3 showed that bilingual competence, particularly in linguistic terms, can be characterised by not only the static breadth of lexical knowledge (i.e., lexicon size) but also by more dynamic dimensions in its depth, including the lexical strength of the connections between each component of a bilingual memory model, the efficiency and automaticity of lexical access, and the completeness of lexical acquisition in terms of the adequacy of information that becomes available to bilinguals when a certain lemma is selected.

However, these findings were largely based on the data collected from bilinguals in monolingual context. Therefore, one might ask whether the same findings can generalise to sentence translation. And even if they can, does it add knowledge to our understanding of bilingualism? A comment that partially responds to these questions was made by Thierry and Wu (2007, p. 12530), "previous studies have made extensive use of cross-language priming or overt translation tasks to compare native- and second-language activation in bilinguals... mixing stimuli from two languages creates an artificial context that necessarily biases the output of behavioural tests toward a bilingual or 'dual-language' activation pattern. For that matter, translation tasks are even more biased because they require conscious access to both languages." Thierry and Wu's (2007) comment is valid, and in fact some doubt has been cast on applying theories or paradigms of bilingualism research in interpreting studies. There is no denying that one has to be cautious in generalising conclusions across tasks. But the challenge to interpreting studies can be defended on the ground of motivation. Note that translation is a cross-culture communication skill and a function coexisting with bilingualism. When bilingualism studies focus on the fundamental knowledge of the organisation and operation of human languages, translation research is interested in finding out what happens if the same group of bilinguals, whose performance in bilingualism studies have helped shape an understanding of bilingual functioning, take part in a translation task. For instance, to the extent that lexical access is language-nonselective in bilinguals' word recognition and production, language interpreting that requires both language comprehension and production is an ideal task that can be used to test the hypotheses of whether language-nonselectivity principle holds, particularly in light of a suggestion that interpreting requires two languages to be constantly, if not equally, active

(D. Green, 1993). Another related question of interest is to what extent nonselective lexical access is automatic in discourse translation? And given that bilinguals often show unbalanced proficiency in their languages, sentence or discourse interpreting could complement other research which uses a correlational/individual difference approach by testing theoretically driven hypothesis regarding interpreters' resource allocation. Specifically, this thesis explores how and at which stage working memory is implicated in discourse interpreting.

2.4 Translation components – A full account of resource use, comprehension, memory, and language production

In previous sections, it was proposed that a satisfactory model of interpreting should at least be able to account for the performance of the interpreters' language proficiency, but above all, it should address a more general question as to whether there are factors that contribute to the variance in their performance, independently of their proficiency.

Since existing theories of interpreting/translation are largely motivated by pedagogical considerations and their specification does not generate testable hypotheses, a full consideration of subcomponents in interpreting has been put forward (section 2.2.1) and will be elaborated corresponding to the questions raised in section 2.2.1. The rationale was that subcomponents of interpreting should ideally be discussed and examined together, as they are interrelated. This wholistic approach is especially crucial when the assessment of task performance includes off-line measures, e.g., recall protocols. And given that participants in interpreting studies and a large proportion of interpreters are potentially late bilinguals or indeed L2 learners (Campbell, 1998), interpreting would demand cognitive resource for supporting processes that are less automatised (see section 2.3.3). In the following sections, the four components of resource, comprehension, memory retention, and language production will be respectively discussed. In the discussion of the three components of comprehension, memory and language production, candidate theories that might help address the research questions will be introduced. In addition to the theoretical background for each translation component, evidence in relevant monolingual, bilingual, and interpreting studies will be reviewed. The theories and evidence reviewed for the three components of translation would then lead to questions that require knowledge of the fourth component, cognitive resource, in finally formulating research questions, and offering methodologies to address these research questions.

2.4.1 Comprehension component in translation

There are two issues that most review articles in translation/interpreting theories have singled out in their discussion (Bajo et al., 2001; Christoffels & De Groot, 2005; De Groot, 1997; P. Padilla et al., 1999). On one hand, the translation procedure can be distinguished by meaning-based and form-based strategies in which form-based strategy involves a transcoding process whereby an SL word can directly map onto its TL translation equivalent. On the other hand, in the majority of translation theories/models reviewed earlier (section 2.2.2), 'meaning' is the central theme in models of various sophistication. One shared feature among these models is their underspecification of what they imply by the 'meaning' that is extracted from an SL discourse. Without the knowledge of the characteristics of the mental representation constructed in the comprehension phase, the theory development of translation and interpreting is difficult to progress, and empirical studies are not theoretically grounded. On the other hand, the distinction between form- vs. meaning-based translation strategy, though useful in conceptualising translation processes, appears oversimplified, in that it does not formalise how this strategy may be applied, and particularly for how long the 'form' of an SL item can be retained in memory in order to be mapped upon in a TL if form-based translation is used. This section will review major comprehension theories and explore whether these theories have a space for comprehension strategies suggested in interpreting theories. Since little is known about the comprehension process in interpreting, this section will start by introducing an influential discourse comprehension theory, and continue by reviewing factors that are known to affect discourse comprehension process and its subsequent memory recall with a focus on the content as opposed to the verbatim form.

Discourse Comprehension

Comprehension is one of the cornerstones of effective and accurate interpreting. Even though it is not possible to find out exactly what level of comprehension is required for interpreting, it is by no means a moot point as has been suggested (Pöchhacker, 2004b) in trying to identify the relations between the characteristics of SL discourse, the retention of parsed material and the ease with which SL discourse content is retrieved during TL production.

One of the most influential and widely adopted discourse comprehension models is the Construction-Integration (CI) Model (Kintsch & Van Dijk, 1978). In the CI model, different processes are engaged in constructing three levels of discourse representation: surface code, text base, and situation model. These processes include word recognition, parsing, inference drawing, and gist extraction. A sentence *John hit the big red ball* can be parsed and broken down into multiple propositions, each consisting of predicates and arguments as follows,

P1: Hit (John, ball)
P2: Is (big, ball)
P3: Is (red, ball)

Hit is the predicate of proposition one (P1), whereas *John* as the subject and *ball* as the object, compose an argument, which in other cases can be an embedded proposition (Kintsch & Van Dijk, 1978). "The arguments of a proposition fulfill different semantic functions, such as agent, object, and goal. Predicates may be realized in the surface structure as verbs, adjectives, adverbs, and sentence connectives" (Kintsch & Van Dijk, 1978, p. 367). When a discourse is composed of multiple sentences, the propositions constitute a hierarchical structure (Kintsch, 1994), also known as a text base or microstructure (Figure 2.33).

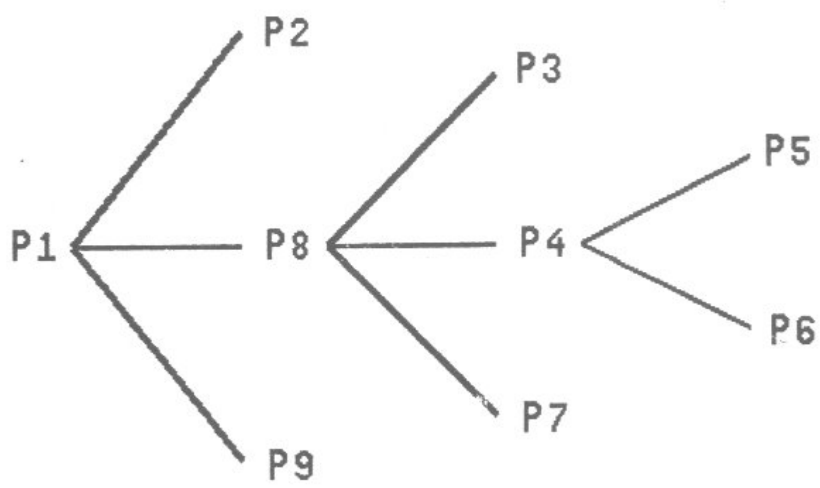


Figure 2.33: A hierarchical structure of propositions derived from a prose. Adapted from Miller (1984, p. 330).

It is suggested that comprehension always involves knowledge use and inference processes (Kintsch & Van Dijk, 1978). The knowledge-based top-down process integrates the local-level microstructure with knowledge stored in long-term memory and creates a macrostructure (Van Dijk & Kintsch, 1983), also known as situation model. When people read sentences for comprehension, they generate three types of representations: the surface code, the textbase, and the situation model (Van Dijk & Kintsch, 1983). The surface code is a representation of the exact words, which is often the form of response required by comprehension or short-term memory studies. However, verbatim sentence memory is short-lived (Sachs, 1967). The textbase is more durable and can be more easily retrieved from memory. "Studies looking at memory for propositional content, paraphrasing ability, and ability to follow directions are aimed at the textbase level" (Radvansky & Copeland, 2004, p. 193).

Language interpreting has been conceived as textbase-focused processing by some interpreting researchers who adopted Kintsch and Van Dijk's (1978) model, e.g., MacKintosh (1985) and Dillinger (1994). The situation model is the representation of the situation described by the text. It is the gist of a discourse. A situation model is created by the *macro-operators* with procedures of "deleting or generalising propositions that are either irrelevant or redundant and by constructing new inferred propositions." (Kintsch & Van Dijk, 1978, p. 372) Importantly, the operation of macro-operators is guided by the schemata that specifies both the schematic categories of the texts (e.g., a research report) and also by what information in the texts is relevant to the macrostructure (Kintsch & Van Dijk, 1978). When a reader has a special purpose, the reading process is likely to be controlled by the purpose set by a task schemata rather than the usual story schemata. In this regard, comprehension for translation necessarily would require a unique schemata that guides interpreters' construction of adequate representations of a discourse under time pressure and under potential influence from the factors that affect general comprehension.

The CI model has been adopted by many researchers of text processing, e.g, Raney, Obeidallah, and Miura (2002) in bilingual sentence processing; MacKintosh (1985) in language interpreting, not only because it is a model that considers and specifies discourse comprehension, memory, and production, but also because it has been simulated computationally (Kintsch & Van Dijk, 1978). In order to computationally model texts and evaluate behaviour performance, Kintsch and Van Dijk (1978) decided to adopt proposition as the unit of concepts that cannot be reduced any further (the same approach is taken in other models such as ACT-R). Van Dijk and Kintsch (1983) documented evidence of the proposition as a parsimonious means of conceptualising meaning or idea units that can be used in measuring recall performance. For different purposes, propositional analysis systems (Kintsch, 1974; Meyer, 1975; A. Turner & Greene, 1977) have been developed on the basis of early linguistic works (Bierwisch, 1969; Fillmore, 1968). Their differences are merely notational (Kintsch, 1994). A. Turner and Greene's (1977) system, which was an extension of Kintsch (1974), was chosen as the tool for the propositional analysis in this thesis, as it offers clear guidelines and examples that make text propositionalisation more likely to be consistent. According to Kintsch and van Dijk, "with some practice, persons using this system arrive at essentially identical propositional representations." (Kintsch & Van Dijk, 1978, p. 367) A. Turner and Greene's (1977) system will be described in detail in Chapter 3.

The CI model is a process model that describes the procedure of discourse comprehension and the algorithm helping to implement it computationally. It is not surprising that there are complicated procedures and many assumptions made. But since the focus of this section is the comprehension component of translation, a summary of the procedure of discourse comprehension and assumptions made in the CI model relevant to comprehension in translation is presented below. The aspects of memory retention, and memory retrieval in the CI model will be introduced later in turn.

- The CI model can be applied to both reading and listening comprehension (Kintsch & Van Dijk, 1978). During discourse comprehension for a text composed of n propositions, part of working memory can buffer a limited size of s propositions. A discourse is thought to be processed in chunks of several propositions at a time.
- Checking argument overlap: Discourse comprehension proceeds in cycles and assumes a resource component in the process. No more than two propositions buffered in working memory will be used to connect new incoming propositions. "When there exists some argument overlap between the input set and the contents of the buffer, the input is accepted as coherent with the previous text" (Kintsch & Van Dijk, 1978, p. 368). Otherwise, a resource-demanding

search in long-term memory or an inference process is initiated, which adds one or more propositions to the text base.

- Processing efficiency and familiarity: Although there exist individual differences in the buffer, the capacity also depends on the automaticity of processes like perceptual decoding, syntactic-semantic analyses and inference generation. The efficiency of the comprehension process is partially determined by the topic familiarity and the complexity of the surface structure (Kintsch & Van Dijk, 1978).
- Probability of recall: "The value p (reproduction probability of propositions) depends on the task demands that govern the comprehension process as well as the later production process" (Kintsch & Van Dijk, 1978, p. 371). For instance, reading a long text for gist comprehension could lead to lower probability of storing propositions individually than reading for immediate recall. The same can be said of reading for summarising vs. reading for immediate recall. When less capacity is required for lower-level processes, more is available for storing individual propositions in memory. A crucial assumption here is that familiarity is related to chunking efficiency during processing and storing.

In the CI model, working memory was assigned the role of buffering two propositions in the cyclical process of comprehension. A multi-component model of working memory is conceived as a workspace for information storing and processing, so when proposition buffering may be the *storing* function of working memory during discourse comprehension, it can be asked whether working memory also plays a role in information *processing* during comprehension. One hypothesis entertains the concept of information binding. In this binding theory, information of different modalities can be bound with the facilitation of long-term memory and then stored in a hypothetical component which is called the episodic buffer (Baddeley, 2000;

Baddeley et al., 2009). Another hypothesis considers that comprehension has become an expert skill of effectively every literate human being. Ericsson and Kintsch (1995) distinguished Long-Term Working Memory (LTWM) that can account for expert skill or expert performance from Short-Term Working Memory (STWM) that has difficulty in accounting for performance outwith the domain of tasks or skills that require rote memory, e.g., patients who had impaired phonological store showed nearly normal immediate prose recall (B. Wilson & Baddeley, 1988). The LTWM and episodic buffer will be discussed in turn in the following sections.

During discourse comprehension, new propositions (nodes in Kintsch & Van Dijk, 1978) and connections are formed through the construction-integration process. The pattern of the complex network is determined by "the nature of the text and the comprehension strategies of the reader" (Kintsch, Patel, & Ericsson, 1999, p. 5). One type of the links thought to be formed between the newly formed nodes and some nodes in long-term memory is through a pre-existing LTWM. LTWM is conceived as a subset of long-term memory. Memory retrieval in LTWM is cue-dependent through short-term working memory. Because memory cues are thought as part of a retrieval structure, as long as cues are linked to long-term memory, retrieval operation is automatic and quick. However, this type of structure is thought possible only in the domains of expert skills or tasks. The approach that links LTWM and comprehension is the proposal that one expert skill shared by literate individuals is language comprehension (Kintsch et al., 1999). One crucial point that is relevant to most discourse memory studies is that the ability to retrieve is incidental to comprehension. If one comprehends a text, a mental structure has been created that can support memory retrieval via LTWM. What is essential for robust LTWM is thought to be "appropriate comprehension strategies, the knowledge, and skills necessary for the use of these strategies" (Kintsch et al., 1999, p. 6). The connectionistic approach in Kintsch and Van Dijk's (1978) CI model is again the central feature in LTWM. In a flashlight metaphor, Kintsch et al. (1999) described how LTWM can be conceptualised.

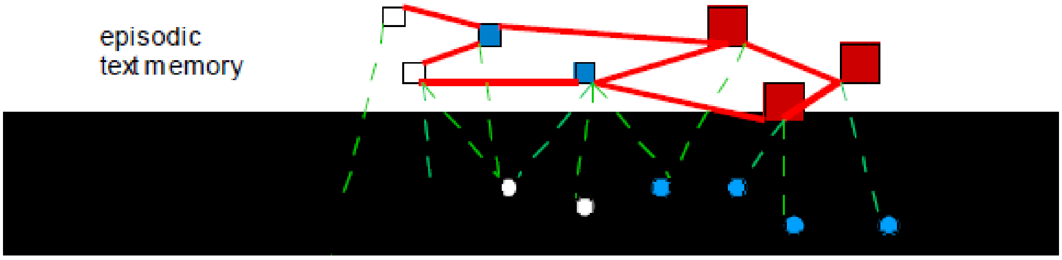


Figure 2.34: A schematic representation of LTWM. Adapted from Kintsch (1999, p. 6).

In Figure 2.34, red squares are three propositions that are still active in short-term working memory and the other squares are not in the focus of attention. But these squares are strategically linked (red thick lines) during comprehension. In the dark area (long-term memory), blue nodes are directly linked to the contents of short-term memory and comprise LTWM. Initially, the LTWM network has relevant as well as irrelevant nodes when the representation is constructed, but after the process of integration, the irrelevant nodes are deactivated while the relevant nodes in long-term memory and short-term memory remain strongly activated (Kintsch et al., 1999).

As for the episodic buffer, it shares at least one hypothesis with LTWM in the domain of sentence memory. Both assume that the units held in working memory are the structures constructed during discourse comprehension, i.e., propositions or a text base. But they have quite different phenomena that they aim to and can account for. Moreover, the differences make their comparison rather tricky. When LTWM conceives the memory for discourse as a retrieval structure that consists of a network of connected nodes in short-term and long-term memory, and it measures memory in terms of number of propositions; the episodic buffer conceives the sentence memory as multi-dimensional chunks, each of which consists of a sequence of words (Baddeley et al., 2009). Therefore sentence memory could be measured in the number of words reproduced in a recall protocol (Baddeley et al., 2009).

While LTWM proposes that the retrieval structure is a by-product of discourse comprehension and is not specifically constructed for the purpose of memory recall, one of the episodic buffer's main concerns is how a sentence is better recalled verbatim than unrelated words. The conceptual space for temporarily held information also differs between LTWM and the episodic buffer. The retrieval structure of LTWM is thought to be an interface between short-term and long-term memory, whereas the episodic buffer is an independent component in working memory. These differences could lead to different predictions for memory performance in a translation task. As memory tasks have been commonly used to assess participants' discourse comprehension, and memory has been suggested as a by-product of sufficient discourse comprehension, it is necessary to know what factors have been shown to affect comprehension.

Factors that influence comprehension

Relative to the underexplored area of working memory in discourse comprehension, a considerable body of research has established potential roles working memory plays in sentence comprehension and sentence memory (Baddeley, 1983; Daneman & Carpenter, 1980; Just & Carpenter, 1992). These roles can be discussed from the perspectives of structural and resource constraints. Regarding structural constraints, sentences with complex structures require a temporary store to hold an intermediate representation of processed material until readers get to the point when it can be integrated with incoming material (Gathercole & Baddeley, 1993; Gibson, 1998). In other instances, working memory is required because sentences are ambiguous (Frazier, 1978, 1995) or they contain a list of content words (Baddeley, Eldridge, & Lewis, 1981). In relation to resource constraints, as briefly mentioned in the paragraph about individual differences in sentence comprehension, low-span readers have been shown to be more susceptible to concurrent load and to complex sentences than high-span readers. Apart from structural and resource constraints, other factors have also been found to affect readers' comprehension. Further discussion is focused on translation-relevant factors: reading purpose and input language.

Comprehension Purposes

Most day-to-day use of language, e.g., conversation, rarely requires repetition of what an interlocutor has just said. But if it is true that the retrieval structure of LTWM is incidental to discourse comprehension (see Van Dijk & Kintsch, 1983, p. 335 for a similar argument; but see Kintsch et al., 1999 for exceptions), it is reasonable to predict that Ericsson and Kintsch's (1995) comprehenders are able to retrieve a substantial amount of information in a conversation, given that a robust retrieval structure has already been constructed during comprehension in the first place. But why and when a comprehender has to pay special attention to memorising the content of a discourse? This appears to be related to motivation and goals. Current evidence has lent some support to the view that readers process and recall texts differently depending on their reason for comprehension (Linderholm & Broek, 2002). In a recent study, college students recalled more idea units in reading for study than reading for entertainment (Broek, Lorch, Linderholm, & Gustafson, 2001). Linderholm and Broek (2002, p. 778) suggested that readers' pattern of cognitive processes during reading affected memory of text information. In addition, readers' recall performance was also constrained by their individual differences in working memory: low-span readers recalled less than high-span readers in the study-for-examination condition, but not study-for-entertainment condition (Linderholm & Broek, 2002). Black (1981) showed that different reading purposes led readers to construct representations at different levels in Kintsch and Van Dijk's (1978) CI model. When the instruction of reading was to rate comprehensibility, readers seemed to merely construct a microstructure in memory, but when the task after reading was free recall, readers appeared to construct all three levels of discourse representation. Similarly, tasks required after listening comprehension also affected recall performance of listeners. J. Rubin (1994) discussed a study by Randall Lund[13] in which the comprehension task and participants' year of college affected their recall of task-specific propositions, the proportion of the macropropositions to the micropropositions recalled, and the number of distortions.

Of direct relevance to this thesis are a series of experiments that investigated interpreters' or bilinguals' reading time for later sentence repetition and translation. Macizo and Bajo (2006) asked whether reading for translation was equal to reading for comprehension. In their first two experiments, they employed Miyake, Carpenter, and Just's (1994) sentence manipulation and showed that translators and bilinguals' reading times were longer for ambiguous words than control words in reading for translation, but not in reading for repetition, suggesting that "reading for repetition and reading for translation differ in the amount of working memory resources they require". (Macizo & Bajo, 2006, p. 10) They went on to explore the loci of this reading time difference in their second series of experiments by manipulating the cognate status of critical words in sentences. Since cognate status has been used to show influence from bilinguals' non-target language when a task was conducted in a monolingual context (T. Dijkstra, Grainger, & Van Heuven, 1999), it was hypothesised that if reading for translation did involve activating a translation equivalent, words that were cognate between languages should show an advantage over words that were not. Macizo and Bajo (2006) confirmed their prediction by showing a cognate advantage in reading for translation but not reading for repetition. One crucial finding that contrasted the language-non-selective activation was that cognate status did not affect reading time in reading for repetition conditions. This finding can be taken to rule out, that simply reading words in

[13]In Rubin's (1994) reference, Lund's paper's status was 'submitted for publication.' But it does not seem to be available anywhere.

a target language (e.g., Spanish) would activate their equivalents in a bilingual's non-intended language (e.g., English) in a sentential context. This finding is also in line with the disappearance of cognate effect when bilinguals read contextually high-constrained sentences (Schwartz & Kroll, 2006; Van Hell & De Groot, 2008).

Different purposes have also been shown to lead to slightly different recall performance. S. Lambert (1983) discovered that verbatim recall was slightly better after participants simultaneously and consecutively interpreted sentences than after they repeated sentences during and after sentence presentation, suggesting that operating in two languages increases rather than diminishes depth of processing. However, professional interpreters produced fewer propositions in translation condition than repetition condition, and within the translation condition their translation showed more errors in simultaneous than in consecutive interpreting (MacKintosh, 1985). This study had another objective, and that was to evaluate Kintsch and Van Dijk's (1978) model of discourse processing. From her observation of interpreters' notes taken in consecutive interpreting, she concluded that "the model appears to fit the processes involved in consecutive interpretation and to provide us with a more discrete view of the various stages" (MacKintosh, 1985, p. 42).

Taken together, in order to fulfil the purpose and requirement of accurate translation, there is a good reason to suggest that interpreters not only construct a representation at a higher level, e.g., situation model, like the participants in Black's (1981) study, they may also need to pay attention to the surface code of SL discourse.

When different purposes lead readers to process discourses differently from the way they normally do, they could entail different resource demand as the examples above showed (Linderholm & Broek, 2002; Macizo & Bajo, 2006). What the differences are in processing for different purposes is an empirical question, but one of the possibilities is that readers have strategies at their disposal. According to Van Dijk and Kintsch (1983), strategies are necessary for four reasons:

1. Language users have limited memory, and, especially, a limited short-term memory.
2. Language users cannot process many different kinds of information at the same time.
3. Production and understanding of utterances is linear, whereas most of the structures the rules pertain to are hierarchical.
4. Production and understanding require not only linguistic or grammatical information, but other information as well, for example, information about the context, episodic memories, knowledge of the world; intentions, plans and goals and so on (Van Dijk & Kintsch, 1983).

Although all of these reasons are relevant since this thesis deals with communication across languages, resource limitation and the linearity problem are two issues central to the topic and research questions. They will be discussed in a later section, grammatical encoding in production. Because language interpreting involves at least two languages, and the pace is usually fast, the factors of input language and time pressure are discussed below.

Input Language

Interpreters either translate from or into their first languages. In practice, an ongoing debate that shows no sign of settlement concerns whether the rule of thumb remains that ideally interpreters should translate into their L1 as the Nairobi UNESCO Declaration of 1976 suggested: "...a translator should, as far as possible, translate into his own mother tongue or into a language of which he or she has a mastery equal to that of his or her mother tongue" (Thelen, 2005, p. 242). Many studies focus on translation directionality at the word level, but studies on the effect of directionality at the sentential level are limited (Chang & Schallert, 2007). The central issue in the debate has been whether people should count on better comprehension ability (S. Williams, 1995) or on natural and less error-prone expression in interpreters' L1 (Donovan, 2004). There is evidence that translating into L1 produced proportionally more propositions than when translating into interpreters' L2 (Tommola & Helevä, 1998). On the other hand, interpreters made more errors in grammatical terms but fewer errors in message accuracy when they translated into their L2 (Lee, 2003). Importantly, difficult texts could lead to information loss when the interpreters have trouble in expressing them in their L2 (Daro, Lambert, & Fabbro, 1996). Chang and Schallert (2007) reported that Chinese/English professional interpreters produced more accurate translation when translating into their L1 (Chinese) than into their L2 (English). But when the interpreters' first language was English, there was no effect of directionality. In terms of grammatical errors, there was an effect of directionality, i.e., participants made mistakes more frequently (errors per minute) in translating into their L2. However, the task in Chang and Schallert (2007) is simultaneous interpreting. Under a great time and resource constraint, strategies have been used by interpreters to ensure quality and coherent interpretation at the cost of information omission due to condensation or generalisation (Janis, 2002). It seems as though these strategies could confound the interpreting accuracy as measured in the proportion of correctly reproduced idea units. Taken together, if directionality effect does exist, how should it be taken into consideration in the discussion of the comprehension component? One approach is to establish whether interpreters process their L1 and L2 differently and discuss how the difference in processing may have an impact on later retrieval in TL production.

In section 2.3.3, evidence has been reviewed suggesting that the conceptual representation from L2 words can be weaker than that from L1 words (Duyck & Brysbaert, 2004), and slower word recognition in L2 could lead to a general slow-down in sentence reading (Segalowitz & Hebert, 1990). Segalowitz and Hebert (1990) suggested that less skilled L2 readers might rely on phonological codes when new input cannot be integrated efficiently with prior information and they could easily overload their working memory. This hypothesis is in line with the hypotheses of Kintsch and Van Dijk's (1978) CI model on the cyclical process, in which overload occurs when new propositions cannot be integrated efficiently as more propositions keep entering the system. Segalowitz and Hebert (1990) predicted that the reliance on phonological codes and working memory overload could lead to the loss of some of the non-phonological information. In addition to slower recognition, Vasos (1983) gave evidence to support the view that L2 words did not activate bilinguals' semantic representations as deeply as L1 words. At the syntactic level, however, sentence processing in bilinguals' L2 does not seem dramatically different from their L1 controls, particularly when bilinguals' proficiency increased. Similar conclusions on listening comprehension were made by Voss (1984).

As languages very often differ in their word orders, word order variation naturally arouses curiosity as to how it affects reading and listening comprehension, especially in a second language. Glisan (1985) hypothesised that L2 sentences with similar word order to that of a listener's L1 should be easier to comprehend, but listeners would experience great difficulty when the word order in L2 sentences happened to be inverted canonical L1 word order. Glisan (1985) capitalised on the great variation in Spanish word order and tested his hypothesis by asking English (L1) learners of Spanish to listen to key Spanish sentences in which word orders were manipulated in the forms of S-V-O, V-S-O, and O-V-S. These key sentences were embedded at different positions in short passages. The dependent measures were listening comprehension questions and multiple choice questions to probe word order memory. There was a control group of Spanish native speakers. As predicted, Spanish controls were overall better than L2 listeners but these controls were not sensitive to the word order manipulation. L2 listeners were significantly better in comprehending S-V-O than the other two types of sentences, with the O-V-S condition producing the poorest performance. There was also an effect of key sentence position. Participants performed significantly more poorly in comprehension questions when key sentences were in the sentence-initial position than when they were in the middle or final position in passages. This effect was taken to support the significance of sentence context in comprehension (H. E. Clark & Clark, 1977). It was thought equally plausible that the better comprehension performance for key sentences at the middle or final position was the result of a recency effect of easier accessibility for the most recently encountered proposition. Glisan (1985) suggested that this finding corroborates the idea that short-term memory is a good predictor for language proficiency and functions in both L1 and L2 (Cook, 1975; D. Harris, 1970). Of direct relevance to the discussion in this section were the results of word order retention. Glisan (1985) revealed that L2 listeners' recognition of word order showed the same pattern as that of their comprehension performance. The absence of difference in the retention performance for V-S-O sentences between L1 controls and L2 listeners led Glisan (1985) to conclude that his results was consistent with the claims made in Sachs (1967) and J. Bransford and Jeffery (1971) that listeners lost the surface form of parsed sentences. The contrast of retention of word order between two groups suggested that capacity for retaining surface form is reduced for L2 sentences. One speculative but interesting suggestion made by Glisan (1985, p. 463) who cited Bever (1970) was that "listeners of S-V-O languages look for the first noun-verb-noun sequence to be an agent, action, and object unless the sequence is marked otherwise. The experimental data support the probability that when listeners hear an atypical structure, they transform it into the basic actor-action-object string." If this hypothesis were operative in comprehending for later translation, one cannot help but wonder whether phrasal word order differences affect translation performance in the production stage if word orders become *transformed* early on during comprehension. This point will be elaborated later in the experiment section. It seems that L2 sentences are not intrinsically more difficult than L1 sentences for bilinguals to process. The ease with which sentences are processed is affected by multiple factors, including lower level factors, e.g., word recognition and word order, and higher level factors such as context. This thesis will capitalise on word order differences between Chinese and English in exploring the relationship between bilinguals' comprehension and production systems.

If bilingual readers process L2 sentences in a similar way to L1 processing, do they construct similar representations for L1 and L2 texts? According to Raney et al. (2002), "readers form different representations when reading in their L1 and L2, but these differences seem to reflect the functioning of language-independent reading processes related to linguistic fluency... the content of representation varies as a function of fluency... fluent bilinguals form similar representations for L1 and L2 text" (Raney et al., 2002, p. 173).

To sum up, the comprehension component is one key aspect in translation as it is inseparable from the memory performance, which is why comprehension has been assessed with recognition or recall tasks. But little is known about how interpreters construct an adequate mental representation during comprehension. By following MacKintosh (1985) and Raney et al. (2002), Kintsch and Van Dijk's (1978) CI model was adapted to explore how representation at different levels can be constructed for interpretive purposes and post-interpretive use. Language interpreting involves memory recall, and therefore, one key issue in studying language interpreting is, and has always been, how memory for discourse is formed, stored, and retrieved. Thus two important hypotheses about memory for sentences were contrasted, the LTWM (Ericsson & Kintsch, 1995; Kintsch et al., 1999) and the episodic buffer of working memory (Baddeley, 2000; Baddeley et al., 2009). The implications of their assumptions for the translation task performance were discussed. By the end of this section, the aspect of how representation of a discourse is formed was covered. In the next section, **Memory Component**, the question raised in the beginning: '*what is left in memory after sentence parsing*?' is discussed. The focus in the section

memory component is the predictions made by LTWM and episodic buffer, on how memory is maintained and retrieved before an interpreter begins delivering the TL version of the to-be-translated discourse.

2.4.2 Memory Component

What is left in memory?

Following from the discussion of discourse and sentence comprehension in the last section, the CI model (Kintsch & Van Dijk, 1978) proposes that the product of comprehension is a retrieval structure, also known as LTWM (Ericsson & Kintsch, 1995). In the episodic buffer account (Baddeley, 2000), sequences of words are chunked through a binding procedure with support from long-term memory and then these chunks are stored in the episodic buffer. But how do different theories help us conceive of a plausible mechanism underlying the process of discourse or sentence translation? Possibly, this can be addressed by asking *what is left activated in memory.* This is because language production in a target language in translation will be based on 1) what is left in memory and 2) what is retrieved eventually from memory.

So what is left in memory? According to Van Dijk and Kintsch (1983), the representation a comprehender constructs depends to a large degree on how much support prior knowledge can offer to the comprehension process. When little background knowledge is available, the discourse input is difficult to be integrated with the corresponding situation model and the representation is largely composed of a text-base. But the representation construction also depends on the task of comprehension. In language interpreting, comprehenders do not comprehend for mere understanding and report the gist, as argued earlier, but they have to deliver accurate messages. The attention that might be required by interpreters for details in sentences or discourse could promote a strategy of focusing on facts learning, thus it could involve more textbase-level than situation-model-level processing. This is not to say that interpreters do not construct a situation model, as there is evidence suggesting that information across sentences has been shown to be integrated in participants' translation (Zwaan, Ericsson, Lally, & Hill, 1998). But the support for textbase-focused processing can also be found, such as in the comment, "studies looking at memory for propositional content, paraphrasing ability, and ability to follow directions are aimed at the textbase level" (Radvansky & Copeland, 2004, p. 193). If it is also the case in comprehension sentences or discourse for translation, it may be safe to assume that the kind of mental representation constructed during comprehension does not differ in terms of the level, according to the perspectives of the CI model and the episodic buffer hypothesis. In both accounts, the representations can be conceived of as chunks of information or a hierarchical structure of propositions. A little inconvenience, however, is that the unit of chunks seems to differ between LTWM and the episodic buffer. The former is thought to be a network of interconnected nodes (Figure 2.34), and each node can be seen as a proposition; whereas the episodic buffer is currently focusing more on verbatim word memory and hints that chunks are sequences of words[14]. Radvansky and Copeland (2004) have found a significant correlation between working memory span and recognition accuracy of 'non-functional' sentences[15]. Radvansky and Copeland (2004) attributed this result to a bigger role of working memory in the textbase- than in the situation-model level processing. Following this line of reasoning, there is a need to focus the level of discourse processing and the role of working memory in a translation task on the textbase level, as it seems to be the level that is more comfortably accommodated by Kintsch and Van Dijk's (1978) CI model and the episodic buffer.

[14]Baddeley et al. (2009) did not assess or define chunking directly. But they presumed that better memory for sentences than unrelated words was a result of more efficient chunking due to the redundancy of words in sentences.

[15]Non-functional sentences were those whose messages do not contribute to integration in discourse comprehension. For instance, when a discourse context is about the weather, a non-functional sentence can be *David was standing next to an old bridge* whereas a functional sentence is something like *David was standing below an old bridge*. The functionality difference is that in the latter, the spatial relationship between David and the bridge allows him to get out of the rain.

If comprehension for translation can be approached from the perspectives of a binding process vs. a text-base construction, do they generate different hypotheses for representation constructed for discourses that vary in their input language and structural variation? The last section reviewed evidence that proficient L2 comprehenders construct similar representations for both L1 and L2 discourse. What distinguishes proficient comprehenders from less proficient ones could be the efficiency of binding or chunking. When comprehending L2 discourse, the support that bilinguals' knowledge base lends to integrating new information with old knowledge is constrained by L2 linguistic and metalinguistic knowledge, and this support is less sufficient than that from a bilingual's L1 knowledge. It is very likely that chunking is less efficient in bilinguals' L2 than in their L1 when given a task of verbatim recall for discourse. But a crucial difference between discourse translation and discourse recall is that translation focuses on the message of the discourse, not the wording, which is itself pointless, as languages differ in their word order most of the time. So what translators chunk or bind is more likely to be a kind of representation at a level higher than words. This level is assumed to be the text-base or proposition in this thesis. Given that the number of recalled chunks appears constant (McNulty, 1966), it would seem to be the density of information which can be packed in each chunk that determines how much information can be encoded. In a discourse recall study, Connor (1984) found that text recall in terms of number of propositions was better in L1 than L2 readers, and found that the difference was largely attributable to L1 readers' better recall for subordinate propositions (propositions lower in a hierarchical structure and which are thought less important). This result is consistent with the prediction made in the CI model that the probability for an atomic proposition to be retrieved is higher than that for proposition like modifiers (Van Dijk & Kintsch, 1983, p. 357). The results of Connor (1984) thus implies that all other things being equal, text presented in L1 can be chunked more efficiently than it would be in L2 by bilingual readers. In this case, it would be expected that translators encode less information after comprehending L2 discourse than after comprehending L1 discourse.

When comprehenders construct a semantic representation for a discourse, one would ask whether some syntactic structures are also encoded during discourse comprehension. In section 2.2.3 and paragraph **Bilingual Memory: What is activated in lexical access?** which gave an overview of Levelt's (1989) model, studies that were reviewed observed a robust effect where recently processed syntactic structure (including comprehension and production) or word order is highly likely to be reused in a subsequent linguistic process. This structural priming effect has been observed within and across languages. One of the explanation for structural priming is that recently processed syntactic structure is still activated, and this higher resting activation compared to an alternative structure increases the probability of its reuse. But beyond sentence level, it is often assumed that little or no syntactic information can leave traces in long-term memory. McKoon, Ratcliff, Ward, and Sproat (1993) proposed that syntactic properties in discourse processing appear to underline the most prominent or salient information, and this highlighted information becomes easily accessible and retrievable in short-term and long-term memory. McKoon et al. (1993) used two sentences to illustrate their hypothesis:

7. Bees are swarming in the garden.
8. The garden is swarming with bees.

In the two sentences, McKoon et al. (1993) followed Rappaport, Laughren, and Levin (1987) and suggested that "an argument may be understood to be more affected by the verb if it is placed in one syntactic position rather than another" (McKoon et al., 1993, p. 595). Therefore, it is more likely that the garden is swarming with bees in sentence 8 than in sentence 7. McKoon et al. (1993) also cited D. Wilson and Sperber (1979) who proposed that "the syntactic positions of propositions order them in terms of importance, and that the more important a proposition, the more relevant it is to the discourse as a whole" (McKoon et al., 1993, p. 596). D. Wilson and Sperber (1979) also suggested a reduction in importance when a proposition is expressed in a modifying phrase rather than in a main clause. For instance, *boring* and *expensive* are thought more salient in sentence 9 than in 10.

9. This book is boring, and it is expensive.
10. This boring book is expensive.

In a series of four experiments, McKoon et al. (1993) were able to provide evidence to support their hypotheses and concluded, 1) syntactic prominence increases the chances that propositions are held in short-term memory; 2) their hypothesis of the crucial role that salient propositions play in discourse comprehension is in line with the CI model, in that important propositions will be selected and held in short-term memory during the cyclical construction of the textbase (see section 2.4.1); 3) the prominent proposition and related concepts could form a compound cue and be retrieved together when recall is required.

The evidence reviewed so far indicates that interpreters would focus their processing at the level of textbase for special task demands. For those interpreters who are largely unbalanced bilinguals, they are likely to be affected by the input language, so it is expected that less information may be encoded for L2 than L1 discourse. As has been mentioned earlier, one needs to distinguish between what is encoded (left in memory) and what can be retrieved from memory. Given that interpreters were expected to reproduce less when translating from L2 than from L1, what will be explored further is whether there are other factors that independently affect the retrieval process. It is probable that the final stage of interpreting, language production, could be another major source of performance variability.

How is discourse memory maintained and retrieved?

One well known characteristic of phonological working memory is its storage function via rehearsal with or without overt articulation (see section 2.4.4). But to maintain the memory of a passage composed of several sentences in the form of phonological code would overload the phonological store. This is one of the reasons that Baddeley (2000) found the revision of the original working memory model necessary. However, even though the newly added component, episodic buffer, was assigned the role of binding of information supplied from long-term memory and other subcomponents of working memory, little is mentioned about how the bound information is attended in the buffer. So far, the mechanism of memory maintenance in the episodic buffer is rather speculative. Baddeley (2000, p. 420) hypothesised that "in the case of prose passage, it presumably involves attending to the structure that has been built in order to represent the passage as part of the process of comprehension." Also "that rehearsal within the buffer is more analogous to continued attention to a particular representation"(Baddeley, 2007, p. 151). Since the focus of the episodic buffer about prose recall is on the lexical level, some kind of chaining is thought to be operative, or alternatively, participants' lexical and phonotactic knowledge may be involved during retrieval, especially in reconstructing memory for decayed item representations (Baddeley, 2007, also see Gathercole, Frankish, Pickering, & Peaker, 1999). Regarding memory retrieval, it was assumed that conscious awareness could be the primary mechanism. Baddeley (2000, p. 421) proposed that "[the episodic buffer] is capable of retrieving information from the store in the form of conscious awareness, of reflecting on that information and, where necessary, manipulating and modifying it."

But when applied to retrieval of discourse that spans across several sentences, difficulties emerge as to how a participant determines where to start and how retrieval carries on. So far our knowledge about prose memory is very limited, but one still needs to decide what hypotheses the revised working memory model would make for language interpreting, which relies considerably on memory. According to Baddeley's (2007) hypothesis, maintenance is attention modulated, and together with evidence showing that memory retrieval is less affected by concurrent attentional tasks (Baddeley, Lewis, Eldridge, & Thomson, 1984; Naveh-Benjamin, Guez, & Marom, 2003), it was tentatively hypothesised here that partial cognitive resource will be allocated to the maintenance of discourse memory that is temporarily stored in the episodic buffer for language production in language interpreting. Therefore, single-resource theory of working memory would predict that when a secondary task taps into the same resource pool, the two tasks could mutually interfere with each other. As a result, one might observe that the factors implicated in each task interact with each other (see section 2.4.4 **Approaches**).

An alternative framework is the LTWM. Regarding the maintenance of the memory of discourse, LTWM assigns working memory the role of holding retrieval cues, instead of memory itself. So the maintenance is thought to be related to "skilled readers' ability to access LTM through retrieval cues held in the active portion of working memory" (Ericsson & Kintsch, 1995, p. 229). Ericsson and Kintsch (1995) also hypothesised that it is not the individual differences between skilled and less skilled readers that determine their comprehension performance, rather, it is "skilled readers' superior comprehension strategies that result in the construction of better retrieval schemata" (Ericsson & Kintsch, 1995, p. 229). The retrieval schemata or retrieval structure is assumed to be the text representations that are generated in comprehension according to the CI model (Van Dijk & Kintsch, 1983; Kintsch & Van Dijk, 1978). Ericsson and Kintsch (1995, p. 230) suggested that the retrieval structure is not always propositional in nature, "imagery may be involved that integrates the text and the reader's domain knowledge and supports and supplements the information given by the text with relevant general knowledge of personal experience". If carefully interpreted, this hypothesis can relate to an argument made by Baddeley (2000) in questioning a pure activated LTM account in explaining how people can process novel ideas by simply activating a representation that may have never been constructed before, e.g., ice-hockey-playing elephant. If Baddeley (2000) is concerned about the lack of a computation component (the hallmark of working memory) in LTWM, perhaps Ericsson and Kintsch (1995) need to clarify whether the imagery involved in integrating texts and reader's domain knowledge can be a satisfactory mechanism of information computation.

The maintenance of memory for a discourse that is represented as a retrieval structure of LTWM is, therefore, not directly related to buffering the content per se, but to ensuring that retrieval cues are in the focus of working memory. Van Dijk and Kintsch (1983) adopted Raaijmakers and Shiffrin's (1981) Search of Associative Memory (SAM model) in positing how discourse memory is retrieved. In the SAM model, discourse memory is a network consisting of nodes that can be word concepts or propositions. The connections of nodes vary in strength. It also requires a probe cue to be held in short-term memory for retrieval from the network. According to Kintsch and Van Dijk (1978, p. 375), subjects' memory for a discourse can leave traces of three kinds: "traces from the previous perceptual and linguistic processes involved in text processing; traces from the comprehension processes, and contextual traces." Once an item is retrieved, it may be added to the probe to continue retrieving other nodes. By conceiving propositions as building blocks for chunking, Van Dijk and Kintsch (1983) proposed that subjects might use a text fragment that forms the beginning of a chunk to retrieve the remainder of the chunk. How effective the fragment as a retrieval cue is depends to some degree on whether the cue falls within a chunk boundary. Van Dijk and Kintsch (1983) showed that subjects' continuation of a cue *This discrepancy reflects not only our...* was largely verbatim: *society's concentration of effort* as compared to the original text *society's concentration of formal educational effort*. But when the cue falls on a chunk boundary, e.g., *As shown primarily by the work of Schaie...*, subjects' response were more like a paraphrase of the original text. According to Kintsch and Van Dijk (1978, p. 375), subjects were likely to "apply rules of inference to the information that is still available to reconstruct irretrievable information". The retrieval was assumed automatic and unconscious (Van Dijk & Kintsch, 1983). When participants cannot retrieve any more, they are required to generate retrieval cues to begin resource-consuming searches (Miller & Kintsch, 1980). This scenario was attributed by Ericsson and Kintsch (1995) to readers' failure to construct a coherent text representation in the first place during comprehension.

So how does LTWM with an understanding of the CI model help to formulate hypotheses for recall in translation? One possibility is that memory retrieval depends largely on the comprehension process, and comprehension in turn, depends on factors outlined in section 2.4.1: language input, purpose, and what is left in memory. In that case, all that is needed for discourse memory retrieval is a robust retrieval structure and effective cues in working memory. Resource is required to maintain retrieval cues in the focus of working memory and to generate retrieval cues in cases when retrieval fails. Since the retrieval is cue-dependent, but every time an item is retrieved, the probe array changes as well (Van Dijk & Kintsch, 1983), i.e., it does not appear that a participant has a list of cues in working memory and uses each to retrieve a single item. If this is the case, following McKoon et al.'s (1993) finding that syntactic structure can highlight the most salient proposition, and that atomic proposition is high in a hierarchical structure of proposition, the most salient proposition may be chosen as a retrieval cue in discourse memory retrieval. Unless a concurrent task utilises working memory, it can be assumed to be solely dedicated to maintaining the retrieval cues during recall.

It appears as if retrieval is assumed to be automatic in LTWM and working memory. But when resources are required for memory maintenance, it would be interesting to know whether an interpreter's limited capacity would be thinned out by a concurrent task – target language production. This question depends on knowledge of both memory and language production. On the one hand, in the literature on memory, little is mentioned about the process which materialises the task by producing responses in recall paradigms (Acheson & MacDonald, 2009). On the other hand, language production literature has been focusing on spontaneous word and sentence production in which retrieval for to-be-expressed message in spontaneous speech has a very different demand on the memory system from that required by the retrieval for scripted speech content in an interpretation task. In the next section, the possibility of an interaction between memory retrieval and language production will be discussed with a special focus on L2 production.

2.4.3 The Component of Target Language Production

A question was raised in section 2.2.1 regarding the interaction between memory and language production in spontaneous speech about a particular topic and in recalling discourse. The answer to the question is important, because the framework that this thesis adopted for the study of interpreting is a language production model that can generate a speech plan, but does not have a component that can store discourse content. One way of characterising these two tasks is asking how a speaker feeds the language production system with ingredients. In spontaneous speech, a speaker starts off by generating a speech plan by elaborating the communicative intention and selecting to-be-expressed information, which can sometimes involve elaborative searches in long-term memory, e.g., describing a route from location A to B (Levelt, 1989, p. 126 for examples). But even when a speaker is engaged in that kind of search, the search is guided by the goal of communication, not by what *must be* included in the utterance.

For interpreters, their source of speech plan comes from the discourse memory stored in working memory or LTWM, depending on which framework one chooses to adopt. When discourse topic is familiar to an interpreter, ample support from long-term memory not only facilitates comprehension, it also helps construct a coherent representation and retrieval structure in LTWM. But when a discourse is unfamiliar or incoherent, the chance that these new propositions become integrated with prior knowledge is smaller and consequently the representation is incoherent and less retrievable. In the framework of working memory, the incoherent representation may be overwritten more readily and is less likely to be consolidated as part of long-term memory. So the difference between spontaneous speech and discourse interpreting perhaps lies in where information is retrieved from. If this is so, it follows that the two tasks would place quite different demands on a speaker and on an interpreter, as the interpretation is expected to be accurate and complete, i.e., the requirement of accurate interpretation precludes interpreters from adding their insights or *new* facts that were not present in the SL discourse – their interpretation has to be faithful. Therefore, the difference between spontaneous speech, which enjoys a higher degree of freedom, and interpreting, which has to limit itself only to what is said, underscores the importance of maintenance of discourse memory in interpreting, which is itself resource demanding.

When the ultimate goal of interpreting is a coherent and accurate TL version of a discourse, an interpreter is expected to preserve causal relation or temporal order of the development of events. One example in section 2.2.3 was a difference between *She became pregnant and she married* and *She married and became pregnant.* Levelt (1989, p. 138) suggested that "For event structures, the natural order is the chronological order of events. Unless the speaker explicitly indicates otherwise, the interlocutor will assume that the order of mention corresponds to chronological order." The problem a speaker has is the linearisation problem. This problem has received some attention in discourse processing. Of direct relevance to language interpreting is the variation of word orders in different languages. The linearisation problems can fall into two categories, one at the discourse level, the other, phrasal level (see section 2.2.3 **The Conceptualiser**).

How does the linearisation problem in the conceptualisation of encoding for production interact with memory retrieval? The argument is illustrated using an example given earlier. Consider the sentence below,

11. *John inadvertently gave an old book to Peter in the library.*

According to Levelt (1989), a speaker starts by retrieving a concept in the stage of conceptualisation. By the principle of incrementalism in conceptualisation and grammatical encoding, grammatical encoding for a concept can start before a speaker has the whole speech plan ready. In the example above, it is possible that an interpreter would retrieve the most salient or prominent proposition (John, Book, Peter, GIVE) according to the review in the last section. Since retrieval has been assumed automatic, while memory maintenance is resource consuming, when an interpreter is encoding (John, Book, Peter, GIVE), he or she could be attending to the rest of the propositions in the episodic buffer according to the working memory model, or this retrieved proposition can be added into the probe array as a new retrieval cue according to the LTWM and CI models. Given that the surface structure of the proposition is ready, the next stage, phonological encoding can begin. What is crucial here is that grammatical encoding is not cost free both in L1 (V. S. Ferreira & Pashler, 2002; Smith & Wheeldon, 2001), and L2 (Poulisse, 1997) due to limitations at all levels of linguistic knowledge and their non-automatised procedures (see section 2.3.3). At the same time when an interpreter is attending to the retrieval cues in LTWM (or chunks of discourse memory in the episodic buffer), the rest of the propositions may be retrieved for encoding. As Chinese preposes verb phrase complementisers and modifiers, there is a problem here as to how incremental the grammatical encoding is, and what the unit is, in planning the speech fragment. There are two possibilities. First, an interpreter encodes the main clause, holds the product of the encoding, and retrieves its modifier/complementiser. Alternatively, he/she retrieves the main clause together with its modifiers during conceptualisation before starting grammatical encoding. Intuitively, it seems cumbersome if the language production system is radically incremental, in that it does not wait until the modifier or complementiser are retrieved before it starts to encode the main clause of a message. This is because the product of this radical incrementalism could result in an awkward word order (11a) compared to the canonical one (11b):

11a. 約翰無意間給彼得一本舊書在圖書館裡(Awkward Chinese word order)
John inadvertently gave Peter an old book (in the library).
11b. 約翰無意間在圖書館裡給彼得一本舊書(Canonical Chinese word order)
John inadvertently (in the library) gave Peter an old book.

If F. Ferreira and Engelhardt's (2006) example of encoding the phrase *my cat terrifies the dog next door* that includes the main clause with its modifier can be thought to illustrate a norm of the unit of grammatical encoding, it may be plausible that a Chinese-English interpreter retrieves the main clause and its modifiers/complementisers at the stage of conceptualisation before their encoding. In this way, the word order can be assigned to each item more efficiently at the stage of grammatical encoding. When memory retrieval is assumed automatic and grammatical encoding in L1 is largely automatised, one could expect that memory retrieval and language production in L2-L1 translation does not interact significantly. However, translation in the opposite direction could be a different story. In L1-L2 interpreting, although the stage of conceptualisation should not differ too much from that in L2-L1 translation, partial resource would have to be allocated to support non-automatised L2 lexical access and grammatical encoding. Therefore, there could be resource competition between the maintenance of memory and grammatical encoding in translating from L1 to L2. Given that structural priming has been thought as the means which the production system exploits to reduce its work load

(Corley & Scheepers, 2002; Pickering & Branigan, 1999; Smith & Wheeldon, 2001), it would be expected that during grammatical encoding in L2, the word order that is shared between L1 and L2 would be easier to be determined than those that are not. In this case, it is possible that the demand for resources is significantly larger in encoding sentence 12a than sentence 12b in preparation for production in English.

Word orders are not shared:
12a. 約翰無意間在圖書館裡給彼得一本舊書
John accidentally gave Peter an old book [in the library].
Word orders are shared:
12b. 在圖書管裡約翰無意間裡給彼得一本舊書
In the library, John accidentally gave Peter an old book.

The consequence of higher demand for grammatical encoding in an interpreter's L2 could be less resource for maintaining the buffered chunks or retrieval cues in working memory. Consequently, it might be expected that key retrieval cues or propositions in working memory are not in the focus of attention, and eventually become irretrievable. In terms of interpreting performance, it might be predicted that sentences that have shared structures between languages are less likely to result in the loss of information than sentences that do not share surface structures. It is also predicted that translation direction and word order interact, i.e., interpreters might be differentially affected by the structural differences depending on whether they are translating into L1 or L2. Interpreters should be less affected by the differences in surface structure between two languages when translating into L1. Alternatively, less resource for memory maintenance does not leads to information loss, but lead to retrieval difficulty, the result of which could be a speech with more pauses or lower fluency in terms of effective syllable per second. This possibility is explored further in the experiment section 4.5.

The discussion above focuses on how conceptualisation and message planning might be influenced by the structural difference between SL and TL. In the following section, the focus is shifted to grammatical encoding in a translation task. In section 2.3.3 **Bilingual Memory: What is activated in lexical access**, lexical activation for single word production has generally been shown to be language-non-specific (but see Costa & Caramazza, 1999; Costa, Miozzo, & Caramazza, 1999 for language-specific account), that is, a concept can spread to activate lemmas in bilinguals' both languages. However, opinions divide on how a TL lemma is selected. D. W. Green (1998) proposed that lemmas are tagged with languages they belong to, but lemmas with incorrect tags would be actively suppressed through a central inhibitory control system. In D. W. Green's (1998) Control-Inhibition (CI) model, inhibition is reactive and proportional to the level of activation of the words that are to be suppressed. The magnitude of inhibition is thought to depend on bilinguals' proficiency in their languages, whereby speaking in a less dominant language entails stronger inhibition on their dominant languages. Another crucial hypothesis of CI model is that inhibition on non-intended language has a consequence of more difficult re-activation for the inhibited language. This hypothesis predicts an asymmetry of switch cost for unbalanced bilinguals in a task that requires switching between two languages. Without resorting to a control mechanism, LaHeij (2005); Poulisse and Bongaerts (1994) suppose that lemmas in the intended language are activated to a larger extent than words in the non-intended language. So even when lexical representations from different languages compete for selection, this competition is resolved by lateral inhibition (also see A. Dijkstra & Van Heuven, 1998; Logan & Cowan, 1984). In this regard, bilinguals may not be good suppressors, but they might be good at maintaining the task goals in their working memory. In the context of lexical access in a translation task, this thesis has no intention or power of testing the two competing hypotheses. More relevant to the thesis, however, is the question of how language dominance could modulate the ease with which participants access the lemma in the intended language. There is evidence suggesting that higher proficient bilinguals showed fewer L1 intrusions than less-proficient bilinguals in L2 production (Poulisse & Bongaerts, 1994). When translating from L1 to L2, less proficient bilinguals might need to exert stronger inhibition on their L1 when they access lemmas in their L2. Of course bilinguals' language dominance should not be construed as static or definite. Depending on bilinguals' language preference of terminology use in particular topics, frequency of language use and recency of using a particular language prior to participating in an experiment, bilinguals can differ in their language dominance. The inhibition control account on lexical access might be extended to the syntactical level, in that interpreters might need to inhibit highly activated structural information given in a SL text during the positional processing in TL grammatical encoding. In both lexical and syntactic levels of bilingual processing, the issues of inhibition control need to be discussed not only in the phase of production, but also the comprehension phase. This is because interpreters appear to access not only the SL lexicon for comprehension, but also the TL lexicon during comprehension for later translation Macizo and Bajo (2006); Ruiz et al. (2008). It was found that word reading times in preparation for later sentence translation were modulated by TL word frequency when SL word frequency was controlled, that is, reading times were longer when TL word frequency was low than when it was high (Ruiz et al., 2008). To the extent that this finding holds, it is possible that participants in Ruiz et al.'s (2008) had accessed lexemes (Jescheniak & Levelt, 1994). In other words, they had selected an appropriate TL equivalent with or without interference from SL word and/or TL candidates. Likewise, participants' longer reading times in incongruent conditions when word order differed between English and Spanish (e.g., beautiful garden vs. garden beautiful) versions of adjectival phrases than in congruent conditions could imply that resources were recruited to access TL syntactic structure or to inhibit interference from SL structure (Ruiz et al., 2008).

Thus far, at least one question in each of the three translation components, discussed above, has been raised. In the comprehension component, Kintsch and Van Dijk's (1978) CI model was used in describing how a discourse is comprehended and a representation is constructed, resulting in a retrieval structure. Two factors were highlighted in the relationship between comprehension and subsequent memory performance. One was input language, and the other comprehension purpose. These factors are particularly relevant to language interpreting, since interpreting typically involves more than two languages and the purpose of comprehension in interpreting is not only constructing a sufficient representation, but also making sure the message can be reproduced with high accuracy and fidelity. The latter would require interpreters to process an SL discourse particularly at a textbase level. As Macizo and Bajo (2006) and Ruiz et al. (2008) have established, participants in self-paced reading studies appeared to access TL properties during SL comprehension. It has also been suggested that processing at the textbase level is very likely to implicate working memory. The questions that can be asked for the comprehension component are whether input language and comprehension purpose affect language interpreting in terms of the memory performance, and whether working memory plays a role in language interpreting.

In the memory component of interpreting, it was highlighted that syntactic information might be used by comprehenders to identify prominent propositions in a discourse and these propositions are better recalled because they are more accessible and they might be part of the retrieval probes when recall is required. Another topic of discussion concerns how discourse is maintained and retrieved from memory. Two hypotheses were made, based on the LTWM and the episodic buffer accounts.

According to the LTWM account, discourse memory could be a retrieval structure that is intrinsic to comprehension. Retrieval for discourse memory has been assumed automatic and effortless in the LTWM account. Although the episodic buffer is a developing concept and cannot offer clear specifications, current understanding suggests that discourse memory might be stored in chunks in the episodic buffer and its maintenance demands attentional capacity. One question that can be asked for the memory component, and for the language production component as well, is whether memory maintenance can be so resource-demanding that an interpreter might perform more poorly when the TL production competes for cognitive resources with memory maintenance.

One subject that these questions can all relate to is the interpreters' cognitive resource allocation in a language interpreting task. Ruiz et al. (2008) provided evidence that reading for comprehension was different from reading for translation. They showed that interpreters appeared to access TL information during SL comprehension, but this process demanded working memory. Although there does not seem to be evidence suggesting a trade-off between memory retrieval and language production, the earlier review on unbalanced bilinguals' linguistic competence suggested that grammatical encoding in bilinguals' L2 is resource-demanding. If memory retrieval and grammatical encoding share one resource pool, one task should show decrement in performance when the other overloads working memory. In order to formalise the research questions of the thesis, an in-depth discussion of the resource component of interpreting is provided in the next section. A brief introduction of the WM model will be given before the approaches and tools that are typically used in studying the role of working memory in language processing are discussed. To conclude, the research questions would be formalised and the approaches to address these questions will be described.

2.4.4 Resource component: Working Memory

Working memory is probably not the only resource that is thought to support interpreting tasks. Nevertheless, language comprehension and production involve interaction between new information and prior knowledge. Although there is an ongoing debate about the relation between working memory and long-term memory (Miyake & Shah, 1999; Ruchkin, Grafman, Cameron, & Berndt, 2003), most memory models agree on the need for a temporary store for linguistic processing. This thesis does not aim to resolve the debate, but Baddeley's (1986) model of working memory is widely used and it seems to be a strong candidate to serve as a point of departure in the discussion of the resource component. As a resource component, it can be understood as a process that temporarily stores and processes the information that is perceived, or retrieved from long-term memory. The following paragraphs briefly introduce the major hypotheses and relevant evidence for major components of working memory that are most relevant to verbal activities.

The original model of working memory (Baddeley & Hitch, 1974a) comprised three components: a limited-capacity attention controller, the central executive (CE), assisted by two slave systems, phonological loop (PL), and visual-spatial sketchpad (VSSP) (Figure 2.35). The evolution and empirical testing of the WM model is ongoing. One of the important changes in working memory theory was that its role as a gateway for information processing between perception and long-term memory was conceptualised as a workspace in which information can be temporarily stored and processed concurrently (Logie, 2003). Recently, an *episodic buffer* was added in the working memory model (Figure 2.36) by Baddeley (2000). Language interpreting resembles comprehension and reproduction of sentences to some degree, therefore, following Roberts and Gibson's (2002) assumption that "remembering sentences and answering questions about them could depend primarily on the STS, the CE, or both," the focus in the discussion of working memory is on the phonological loop, central executive, and episodic buffer.

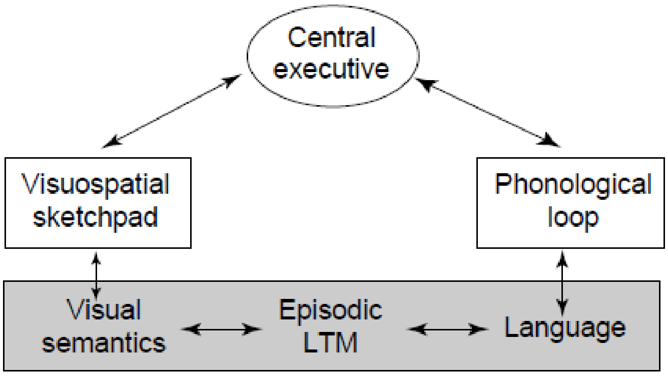


Figure 2.35: The original model of tripartite working memory. Adapted from Baddeley (2000, p. 418)

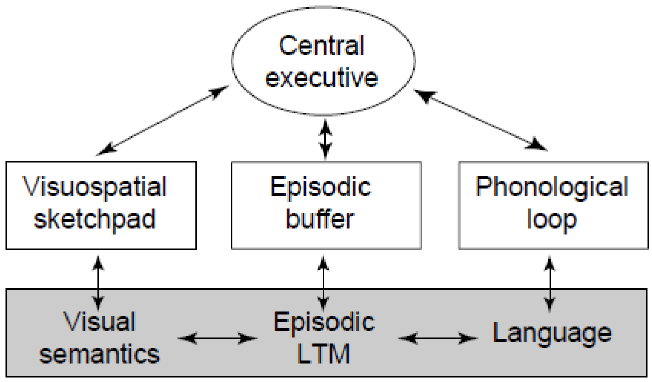


Figure 2.36: A revised model of working memory. Adapted from Baddeley (2000, p. 421)

The Phonological Loop

The phonological loop is assumed to have two components, a phonological store and an articulatory rehearsal system. The phonological store can hold information temporarily, but information can decay over time unless the phonological rehearsal mechanism activates subvocal rehearsal to refresh the speech-based information. Available evidence suggests that the phonological loop has a significant role in the acquisition of native and foreign languages (Baddeley, Papagno, & Vallar, 1988; Gathercole & Baddeley, 1993; Papagno, Valentine, & Baddeley, 1991; Service, 1992), but its role in day-to-day language use seems less important. Vallar and Baddeley (1987) showed that patients with an impaired phonological store had trouble in sentence comprehension only when sentences were extremely long or complex, suggesting that the phonological store might serve as a backup system in comprehension. The limited capacity of the phonological store has been borne out by the identification of several effects associating list memory with the characteristics of to-be-remembered items, e.g., word length and phonological similarity (Baddeley, Thomson, & Buchanan, 1975; Conrad & Hull, 1964), with the capacity to generate speech-motor programmes (Baddeley & Wilson, 1985; Caplan & Waters, 1995), and language familiarity (Thorn, Gathercole, & Frankish, 2002). The findings in vocabulary acquisition and first language superiority in serial recall, in particular, underlined the interactive nature and a potential interface between working memory and long-term memory in certain tasks.

The Central Executive

The central executive has been construed as an attentional controller or as a capacity-limited attentional resource (Robbins et al., 1996), which is able to divide attention for concurrent tasks (Baddeley, Bressi, Della Sala, Logie, & Spinnler, 1991; R. H., G., S., & A. D., 2004), as being responsible for switching attention between tasks (Allport, Styles, & Hsieh, 1994; Baddeley, Chincotta, & Adlam, 2001), and finally as being able to form an interface between long-term memory and subsystems of working memory (Baddeley, 1996). When the CE was thought to be a capacity-limited resource, it was predicted that a concurrent secondary task should disrupt the processing of a primary task. This appeared to be the case (Robbins et al., 1996), but this principle did not apply to all complex tasks. Baddeley (2002, p. 90) suggested that "although concurrent load had a clear effect on learning, it had little influence on recall accuracy". The account of divided-attention function for CE was weakened for minimal or absence of interference introduced by a secondary task on the performance of a primary task (Duff & Logie, 2001; Cocchini, Logie, Della Sala, MacPherson, & Baddeley, 2002). The role of the CE in forming an interface between long-term memory and subsystems was called into question when Baddeley et al. (2009) reported that information binding in a sentence recall task did not seem to require attention.

The Episodic Buffer

As mentioned above, the phonological system can temporarily hold information such as a list of digits or words, and it has been suggested as an emergency back-up in comprehension (Gathercole & Baddeley, 1993). Whereas people's memory for unrelated words is extremely limited, their memory for words that can form meaningful sequences has been repeatedly shown to be well beyond the limit of the famous 'seven plus or minus two'. This was true of the normal population and surprisingly so of amnesic patients (B. Wilson & Baddeley, 1988). Where and how were these words stored? The digit span of the normal population is reduced when people are asked to repeatedly articulate an irrelevant word such as 'the the the...' (known as articulatory suppression). This technique is thought to block subvocal rehearsal. However, even with articulatory suppression, digit span does not drop to zero, raising the question as to where or how residual storage is achieved (Baddeley, Lewis, & Vallar, 1984). Patients who show severely impaired digit span (e.g., 1 or 2 digits) following brain damage, performed better when the digits were presented visually (3 or 4 digits) (Shallice & Warrington, 1974). How did they do this? A study that revealed a visual similarity effect (Logie, Della Sala, Wynn, & Baddeley, 2000) in immediate written recall of visually presented letters and words suggests that there is a visual code people can use for encoding visually presented verbal sequences in addition to a phonological code. This could be addressed by assuming the use of the visuo-spatial store that was part of the original Baddeley and Hitch model. However, Saito, Logie, Morita, and Law (2008) demonstrated that both phonological and visual codes could be used with the same material, raising the question as to how these modality-specific codes are integrated. Saito et al.

(2008) showed independent effects of visual similarity and phonological similarity suggesting that both visual and phonological codes can be used with the same material. Moreover, Baddeley (2000) noted that patients who are densely amnesic forget material within a few minutes of it being presented, nevertheless, can hold a conversation, and so can retain the semantic content of speech over short periods of time. The problem, however, is that the original working memory model did not have a component where the information from two subsystems can be integrated, as the central executive was not assigned a capacity for storing to-be-recalled items. Thus a need for a further working memory component arose.

Baddeley proposed, “the episodic buffer is assumed to be a capacity-limited storage system that is capable of integrating information from a variety of sources. It is assumed to be controlled by the central executive. The buffer is episodic in the sense that it holds episodes whereby information is integrated across space and potentially extended across time” (Baddeley, 2000, p. 421). As regards prose memory, it was proposed that material might be chunked into smaller units and a general form of rehearsal for these units could involve sequential attention to the to-be-recalled units. These units were assumed to be “the structure that has been built in order to represent the passage as part of the process of comprehension” (Baddeley, 2000, p. 420). In a recent study Baddeley et al. (2009) used dual-task techniques to test one hypothesis arising from the revised working memory model. The research question was whether chunking in short-term memory for verbal material depended on attentional resources. They started by assuming that the superiority of recall for sentences over unrelated words was a result of more efficient binding, so that each chunk consisted of more words for sentences than for word lists. If binding taxed limited CE resources, a secondary task that relies on CE should remove or reduce the sentence superiority effect. It turned out that the concurrent secondary task impaired recall, but it did not remove or reduce the sentence superiority effect. Baddeley et al. (2009) concluded that the processes involved in chunking occur automatically, and this was taken as evidence that, contrary to Baddeley's (2000) original hypothesis, the CE does not seem to have a major role in binding of sentential material. For a task like language interpreting which could rely heavily on discourse memory when memory aid is not allowed, the episodic buffer naturally has become a candidate component that can be used to support the demanding task (Christoffels, 2006; F. Padilla, Bajo, & Macizo, 2005).

Working memory in linguistic processing – Tools and approaches

Tools

In order to study the relationship between working memory and linguistic processes, one must have the tools and know how to use them. The tasks available may be divided into short-term memory span and complex span tasks. Short-term memory span, e.g., digit span and word span, can be used to index subjects' phonological working memory. These *simple* span measures, however, do not capture the functional importance of working memory that can process items temporarily held in its short-term stores. Thus complex spans were devised to capture the capacity of dual-tasking: storing information for later use and at the same time processing other information. Two widely applied measures of this kind are the reading span (Conway et al., 2005; Daneman & Carpenter, 1980) and operation span (M. L. Turner & Engle, 1989). One less widely used, but relevant complex span is speaking span (Daneman & Green, 1986).

In Daneman and Carpenter's (1980) reading span task, subjects read sets of sentences with the set size starting from two to six sentences. There are typically three trials for each set size. After reading each sentence, participants have to answer a yes/no comprehension question and memorise the last word of each sentence. After reading all sentences, they need to recall as many of the sentence-final words as possible in the order in which the sentences were presented. The task proceeds until subjects fail to recall all items in three out of three trials. For instance, if a subject recalled all five words for sentences the size of five in two out of three trials, but only recalled three words in one of the trials, the task would terminate and the reading span of this subject is scored as five. Another commonly used measure as a working memory index is the total number of trials in which words are correctly recalled (abbreviated as 'total trials' by Waters & Caplan, 1996b). Although Daneman and Carpenter (1980) did not monitor subjects' accuracy in verifying questions, it was possible to do so (Waters & Caplan, 1996b), and in fact, participants' response time, verification accuracy, and span were all recorded in most experiments of this thesis. In so doing, a composite score calculated on the basis of z-score of each parameter takes consideration of storage and processing, and it was possible to show whether there was a speed-accuracy trade-off. If there is no trade-off between reading speed, verification accuracy, and recall, this could imply that processing and storage are independent (Duff & Logie, 2001; Friedman & Miyake, 2004).

Another increasingly popular tool is Conway et al.'s (2005) automated reading span task, which has been recently translated into different language versions [16]. A major difference between Daneman and Carpenter's (1980) and Conway et al.'s (2005) reading span measures is the to-be-recalled item. In Conway et al. (2005), items were not part of test sentences. Participants read sentences, make plausibility judgements and then memorise each letter that appears after the judgement is made. This design presumably avoids a gist effect whereby participants reconstruct words on the basis of the gist of sentences (Conway et al., 2005).

The operation span was developed on the hypothesis that any highly demanding task would necessarily engage the processing function of working memory. In short, working memory capacity was assumed task-independent (Conway et al., 2005). M. L. Turner and Engle (1989) replaced sentences with simple addition or subtraction problems, complete with correct or incorrect answers. These problems and

[16] Currently there are French, Turkish, Chinese, Dutch, Portuguese, and German versions available (Engle, 2009).

simple multiplication or division problems follow each other in a list. Participants read each arithmetic problem aloud and verify whether the answer is correct. After which, a word is presented visually for later recall. The scoring is the same as that for reading span. In the latest version, Conway et al. (2005) stopped the tradition of presenting sentences/problems in the ascending order of set size – they randomised the order. They justified this decision by suggesting that randomisation can reduce the chance for participants to develop strategies as they can no longer anticipate the the number of sentences in an upcoming trial.

The speaking span was originally devised to capture the capacity of concurrent storage and processing of working memory in the domain of speech production (Daneman & Green, 1986). It was also used to test the question of whether working memory capacity can distinguish skilled and unskilled speakers (Daneman, 1991) as they assumed that speaking involves a highly complex and skilful coordinating of processing and storage requirements. "Speakers must plan what to say and temporarily store the plans until ready to execute them as words, phrases, and sentences. Moreover, at any instant, individuals may be planning what to say next while concurrently executing what was planned moments earlier." (Daneman, 1991, p. 446) Typically, participants are presented with increasingly larger sets of unrelated words, one word per second. After each set, they are required to generate one sentence using each word once. Similar to the reading span, there can be two scores. One is the strict span, which is the maximal set size of words that are all used in generating a semantically plausible and grammatically acceptable sentence. The other score is the total number of trials, in each of which, words in a set are all used in generating a sentence. The latter score has become more popular than the former type of score as it approximates a continuous variable. This measure was used by Hartsuiker and Barkhuysen (2006) and one experiment in this thesis. Hartsuiker and Barkhuysen (2006) showed that a preload of 3-word list had an impact on low-span participants who generated proportionally more subject-verb agreement errors under load than in no-load conditions. Hartsuiker and Barkhuysen (2006) thus joined Garrett (1982) and V. S. Ferreira and Pashler (2002) in suggesting that sentence production is constrained by verbal working memory.

Approaches

When it comes to applying these measures to study the relations between working memory and linguistic processing, generally there are two approaches: correlational and individual differences (Roberts & Gibson, 2002). Typically, in correlational studies (e.g., Waters & Caplan, 1996b), a battery of memory tests and tasks in focus would be administered and a factor analysis would be conducted to show what variables are most relevant to the task of interest. Waters and Caplan (1996b) reported a significant correlation between a working memory measure (total trials) and sentence comprehension. But they further demonstrated that when both processing and storage were taken into consideration in determining how working memory was indexed, the correlation coefficients between their composite z-scores of working memory and the comprehension measure were reliable and larger than using the measure that only indexed the storage function of working memory.

By using a composite score based on three working memory span measures, Roberts and Gibson (2002) established significant correlations between probed sentence memory [17] with n-back [18] and with complex memory span. Another example relevant to this thesis is the finding in Daneman's (1991) correlational study that speaking span positively correlated with verbal fluency in terms of speech rate but negatively with spoonerisms.

[17] Participants did not engage in free recall of heard sentences. Instead, they were required to answer probe questions regarding one of the clauses in the sentence. For instance, a probe question may be *What did the barber do?* following the sentence *The barber lectured the sailor who hit the singer who worked in the jazz club.*

[18] Participants are typically presented with a list of digits or words on a screen, one at a time. After the presentation, they need to respond to a probe by indicating whether this probe was the 'n' last item they saw in the list. For instance, a 2-back memory task requires participants to indicate whether a probe was the second last item in the list they saw.

The correlational approach has also been extended to interpreting studies. Christoffels et al. (2003) took measures of word translation time, picture naming, reading span in English and in Dutch (L1), and digit span in Dutch. They identified significant correlations between the factors of word translation and picture naming latency with measures of simultaneous interpreting. Digit span and reading span were also significantly correlated with simultaneous interpreting, in terms of ratings of how well SL discourse was translated.

Another approach involves dividing participants into high-, medium-, and low-span groups on the basis of their working memory measures, and then comparing their task performance, e.g., in reading comprehension. A temporary ambiguity in reading a fragment of a sentence *The experienced soldiers warned...* arises when readers reach the verb *warned* as it can introduce a verb phrase (see sentence 13 below) or alternatively, it can introduce a reduced relative clause [19] (sentence 14 below).

13. *The soldiers warned about the dangers before the midnight raid.*
14. *The soldiers warned about the dangers conducted the midnight raid.*

MacDonald et al. (1992) asked whether individual differences in working memory measured by reading span constrained syntactic parsing. They hypothesised that multiple representations were constructed initially for any ambiguous phrase, but the individual's capacity determines how soon the unfavoured representations are discarded or became inactive. They predicted that high-span readers can buffer multiple representations and for longer than low-span readers could. But this came with a cost. Thereby, high-span readers' reading time in an ambiguous region would be longer than their low-span counterparts, as low-span readers might hold only the chosen representation or they might give up alternative reading very soon. Their

[19] A reduced relative clause does not have a relative pronoun and often gives rise to ambiguity or garden-path effects (Trueswell, Michael, Tanenhaus, & Garnsey, 1994).

results confirmed the predictions and they concluded that "working memory is a major determinant of whether multiple representations are maintained" (MacDonald et al., 1992, p. 87).

Correlational and individual difference approaches have been widely used to explore the relation between working memory capacity and the performance of tasks of interest. But *how* working memory may be utilised in language processing is an entirely different question. More specifically, people asked whether there is a specialised working memory resource (separate-sentence-interpretation-resource theory, i.e., SSIR theory) underlying representation construction during comprehension independent of a general-purpose resource(Waters & Caplan, 1996a), or is there only one all-purpose resource (Single Resource Theory, i.e., SR theory) that supports both comprehension and the operations that utilise the constructed representation for other purposes, e.g., answering questions or recalling sentences (Just & Carpenter, 1992).

SR and SSIR theories have different predictions for performance in a dual-task paradigm. Caplan and Waters (1999) summarised the two approaches. With the individual difference approach, given that a complex sentence taxed working memory, SR theory predicts that working memory span would interact with sentence complexity, i.e., low-span readers would be affected by the complexity more than high-span readers. In contrast, the SSIR would predict only a main effect of complexity, but not the interaction. The other approach is used to investigate whether a secondary verbally mediated task (e.g., digit memory) interferes with sentence comprehension. The SR theory assumes that the two tasks share the same resource pool, and therefore, it predicts an interaction between digit load and sentence complexity, whereas the SSIR does not. A tacit assumption has been made in this approach that this dual-task involves the central executive. One example of the latter approach compared self-paced listening times in a comprehension task that crossed syntactic complexity and digit load (Waters, Caplan, & Yampolsky, 2003). They showed main effects of the memory load and complexity, but these factors did not interact, suggesting that "working memory system used for assigning syntactic structure is separate from that measured by standard working memory tasks" (Waters et al., 2003, p. 88). They also reported a load effect on the off-line measures of digit recall and plausibility judgement reaction time: participants made proportionally more errors in high-load than low-load conditions and they were slower in load than no-load condition in plausibility judgements. Using a different paradigm, Duff and Logie (1999), Duff and Logie (2001), and Logie and Duff (2007) measured processing span without memory load, and then measured word span with no processing load. They found that processing span and word span were unaffected when processing and memory were combined in a working memory span paradigm. This suggested that processing and memory rely on separate resources (see also Daneman & Hannon, 2007).

The studies mentioned above demonstrated how dual-task techniques have shed light on the debate as to whether working memory is best understood as a unified construct (SR theory) or it has developed specialities to meet various task demands (SSIR theory). Similar ideas can be extended to explore the questions put forward earlier: at which stage and how working memory is involved in language interpreting. Because language interpreting, irrespective its mode, involves stages of SL comoprehension and TL production, selected recent studies in language comprehension and production are reviewed in the next section in order to inform predictions that can be made in the two series of experiments in this thesis.

Except balanced or early bilinguals, interpreters who are late bilinguals might rely on a less stable and incomplete system (J. Johnson, Shenkman, Newport, & Medin, 1996) in L2 processing. It is less stable because of the influence of bilinguals' L1 (A. Hernandez, Li, & MacWhinney, 2005). The processes are implemented differently from it is for natives perhaps because maturational changes could lead to the attenuation of the procedural, but the enhancement of declarative memory systems, hence L2 grammatical processing is thought to rely on fact-based rather than rule-based computation (Ullman, 2001, 2004). L2 processing is also thought to be less automatic and more resource-demanding than L1 processing (McDonald, 2006).

One approach to the question as to whether and how working memory is involved in interpreting is testing the hypothesis of linguistic transformation. The linguistic transformation can be conceived as a process where interpreters map messages from a SL to a TL. If this process requires attentional resources and competes with other ongoing operations, it can create a dual-task condition. In a study using mental-addition as a secondary task, Jou and Harris (1992) found that native English speakers' recall for short passages and the pauses in the passage reproduction were a function of attention (full vs. divided). The divided-attention condition led to more omissions of story content measured by idea units and more pauses. A separate analysis established an interaction between attention condition and pause types, with more within-clause pauses than between-clause pauses in divided-attention condition. Whereas in full-attention condition, there were more between-clause pauses than within-clause pauses. This interaction suggests that completing a clause is attention demanding.

There was also a trade-off between mental addition task and story recall, suggesting that participants could be switching between tasks. Jou and Harris (1992, p. 303) concluded that "retrieving the semantic units from memory while speaking required attentional resources. The mental arithmetic task substantially suppressed the message construction process at the conceptual level." Jou and Harris (1992) also documented more retracing (restarting already uttered elements) and more number of fragments (unfinished constructions) in divided-attention than full-attention conditions, suggesting that constructing a full sentence requires sustained attention and perhaps memory as well. There were other signs of deterioration in production in divided-attention condition, such as softer voice, poorer organisation of information, less distinct pronunciation, and less accurate lexical items (e.g., things, stuff). Recently, similar questions were addressed by using different paradigms from that of Jou and Harris (1992). Oomen and Postma (2001) asked participants to describe cartoon strip in full- and divided-attention conditions and compared the frequency of pauses and repititions in their production. Compared to the story retell task in Jou and Harris (1992), it was thought that story-telling is more demanding, as it requires conceptualisation of the stories in cartoon strips. Therefore it should be more affected than a story-retelling would by a secondary task, which was a tactile figure recognition task. Consistent with the findings in Jou and Harris (1992), Oomen and Postma (2001) found that the dual-task condition led to higher frequencies of production of repetitions and filled pauses. We however have to note that story retelling differ from cartoon description on the demanding of memory retention. The to-be-retold material in Jou and Harris (1992) was auditorilly presented and disappeared afterwards. Whereas the cartoon strips in Oomen and Postma (2001) were constantly accessible during the picture description task. So the loci where the dual-task had impact could be different. In Jou and Harris's (1992) study, participants' attentional resources could have been invested in memory retrieval while participants in Oomen and Postma (2001) might have paid more attention to the conceptualisation in picture description. Speculative at the moment, it is possible that the load could be higher in story retelling since it involves memory retrieval and also conceptualisation, whereas the picture description does not have the element of memory retrieval (episodic memory).

Apart from conceptualisation, there is also evidence suggesting that cognitive resource limitation constrains grammatical encoding (Fayol, Largy, & Lemaire, 1994) (see also K. Bock & Cutting, 1992). Hartsuiker and Barkhuysen (2006) tested two hypotheses regarding the resource utilisation in syntactic encoding. Resource-constrained hypothesis predicts that a secondary task that competes with grammatical encoding would lead to errors, whereas resource-free hypothesis predicts that a secondary task makes no difference. In a fragment completion task that manipulated the cognitive load (no load vs. 3-word preload) and the congruence between grammatical number and conceptual number of the sentence head noun (match vs. mistmatch), it was predicted that participants would make more subject-verb agreement errors in the load than no-load conditions. By taking individual's speaking span, it was also predicted that participants with low-span would be more affected by the load than their hi-span counterparts. These predictions were confirmed by the main effects of load and congruence. Memory span interacted with load, so only low-span participants were affected by load. These results were taken to support the resource-constrain hypothesis.

When subject-verb agreement errors have been shown to occur among native speakers due to resource constraints (Hartsuiker, Kolk, & Huinck, 1999; Hartsuiker & Barkhuysen, 2006), there are strong reasons to predict that bilinguals might be more sensitive than L1 spearkers to grammatical and conceptual numbers (Nicol, Teller, & Greth, 2001; Nicol & Greth, 2003; Jiang, 2004), because L2 processing is thought to be less automatically (Segalowitz & Hulstijn, 2005) and more demanding (Miyake & Friedman, 1998; Michael & Gollan, 2005). Hoshino, Dussias, and Kroll (2010) compared proficient and less proficient English-Spanish bilinguals on fragment completion task, in which grammatical number and conceptual number of head nouns were orthogonally manipulated. They found that proficient bilinguals were able to access both proficient and less proficient bilinguals were sensitive to grammatical number information. However, when proficient bilinguals showed sensitivity to conceptual number information on-line, less proficient bilinguals were able to access the conceptual number when it was an off-line task. If the transformation occurs during production phase in interpreting, assuming that a vertical translation strategy is employed, it is possible that interpreters' resources are shared by at least two processes. One is the memory retrieval and the other is grammatical encoding in a TL. Then the findings reviewed above could have direct implications for the directionality of translation. That is, speech planning for L2 production could be so demanding for less proficient bilinguals, hence content information might be compromised at the expense of resource-consuming L2 grammatical encoding. Alternatively, disfluent speech could be produced as a result of more resource being invested in memory retrieval.

Another possibility is that interpreters allocate their resource during comprhension depending on the tasks at hand (Titone, Prentice, & Wingfield, 2000; Levy, 2008). In cases when interpreters not only comprehended SL text but also accessed TL information for later translation (Macizo & Bajo, 2006; Ruiz et al., 2008), it could create a dual-task condition too, depending on how automatic TL information activation is. When interpreters employ horizontal translation but TL activation is not automatic, average reading time can slow down. Again, the ease with which interpreters access the information of a TL would depend on their language proficiency.

The next section formalises the research questions and describes the rationale for the methodologies used in two series of experiments.

2.4.5 Research questions and approaches

This thesis was motivated by a curiosity as to whether and how cognitive resources are deployed in the process of language interpreting. The formulation of research questions was achieved by raising questions in the four components that were thought central to operationalisation of what translation really involves. As most translation and interpreting theories or models were not suitable to address the questions raised, a bilingual production model based on Levelt (1989) was adopted as a framework, and each of the four translation components, i.e., comprehension, memory, grammatical encoding, and resource, was elaborated. The discussion of each was guided by a consideration of the potential impact each component might have on the target language production.

Research Question 1

In the component of language production, a potential linearity problem in language interpreting was identified, i.e., a retrieved proposition cannot be readily mapped onto the surface form of a target language. The research question is whether the differences of surface structure between languages resulted in the difference of interpretation performance, in terms of the percentage of correctly translated propositions.

The locus of linearity problem in language interpreting was assumed to occur during grammatical encoding. It was proposed that positional processing that sequences word order may be primed by the word order of source language discourse. Since grammatical encoding in L2 is assumed to be resource demanding, translating into interpreters' L2 should be more sensitive to the linear relationship between retrieved concepts and their sequencing than translating into the interpreters' L1. Therefore, it was hypothesised that it was significantly more demanding in encoding non-linear than linear constituents in interpreters' L2, but there should not be any effect of linearity in L1. As memory maintenance requires attentional resources, which are limited in capacity, it was hypothesised that the more resource language production demands, the less becomes available for the maintenance of memory of to-be-translated discourse in working memory, and consequently information could become irretrievable. So overall, one might observe a main effect of translation direction: L2-L1 is better than L1-L2 and an effect of linearity: a linear relation between retrieved proposition and canonical word order of a target language would result in better performance than a non-linear relation would. And finally the two factors might interact, so that translating into interpreters' L2 could be significantly more affected by the linearity issue. In order to create linear and non-linear relationships between an SL discourse and its target translation, the types of sentences that reliably differ in their word orders between Chinese and English need to be identified. The passages selected for testing fell in two categories. One contained sentences whose phrasal word orders were congruent between English and Chinese, creating a linear relation between SL and TL translation. The other category contained sentences whose phrasal word orders were incongruent between Chinese and English, creating a non-linear relation between SL and TL translation. In the remainder of this thesis, these passage and sentence types are described as congruent or incongruent.

Research Question 2

The second research question further explored the finding by Ruiz et al. (2008), asking whether and how working memory is implicated in language interpreting. Specifically, if the results in Chinese-English sentence interpreting replicated the findings of Ruiz et al. (2008) and suggested that parallel translation could implicate working memory, then the following question would be which working memory function was involved.

Addressing this question requires one of the approaches introduced in the discussion of the resource component of interpreting. As the focus of the research question concerns not only whether working memory is implicated but also how it might be utilised in language interpreting, it was decided to employ the dual-task paradigm. When taking an approach typically used to test the hypothesis of working memory as a unified construct or separate task-specific capacities in studying language interpreting, a few assumptions have to be made. First, it was assumed that working memory is a unified construct with limited capacity. By following the SR account (Just & Carpenter, 1992), participants' performance can show decrement when they were engaged in two concurrent tasks, both verbally mediated. In the context of language interpreting, one of the concurrent tasks would be the hypothesised process that subserves parallel translation. Effectively, nothing is known about this process, but if Ruiz et al. (2008) were right in suggesting that parallel translation involved searches for syntactic structures, then the component of working memory responsible for processing might be used to support parallel translation. In this case, when participants were given a secondary task that is resource-demanding, the participants' process might slow down. By taking the dual-task approach (Ruiz et al., 2008; Waters et al., 2003), digit preload was used as a secondary task while the primary task was reading for translation. A prediction made for this dual-task design was that participants' reading time would be slower when both digit preload and the process that subserves parallel translation shared the same resource of working memory. Moreover, there might also be an interaction between the digit preload and word order congruency: with a higher load, parallel translation should lead to much longer processing time for incongruent sentences than congruent sentences.

Chapter 3 presents a series of four experiments addressing the first research question: whether grammatical encoding in a TL can be so resource-demanding that translating discourses that contained incongruent sentences could result in higher proportion of information loss than when sentences were congruent in word order between Chinese and English. Chapter 4 reports four experiments conducted to address the second research question regarding whether and how working memory might be involved in sentence interpreting.

Chapter 3

Discourse Interpreting

3.1 Introduction

In this chapter data from four experiments is presented to explore the research question, whether working memory is involved in discourse interpreting. As the four experiments used the same methodology and material, initially the general background information of the methodology is described. This includes 1) discourse interpreting; 2) word order differences between Chinese and English; 3) material generation; and 4) scoring procedure.

3.1.1 Background

Two main characteristics of the task in these experiments led to the name, discourse interpreting. First, the word 'discourse' distinguishes itself from prose, that may conjure up an association with a particular written literary work. In this thesis, discourse means written or spoken communication. Although discourse can include written materials, which are more often static objects of analysis, the focus of discourse in the study of language interpreting was placed on its function of interaction: it is concerned not only with what information is in a discourse but also how the information is mapped onto the surface form of different languages. Second, by adopting the term ‘discourse’, it could not be confused with typical ‘consecutive’ interpreting in the ways they are practised. Consecutive interpreting usually allows interpreters to take notes while listening to a source-language speech. Since the project set out to examine how structural differences between languages affect interpreting performance that depends on the combination of linguistic skills and memory, it was decided not to include note-taking. As the configuration and requirement of the task resembles free recall in many ways with respect to comprehension and language production, one key element in the experimental design was how the responses were scored. The first research question of the thesis was whether word order differences between English and Chinese affect interpreters’ performance in terms of the proportion of propositions reproduced in the TL utterances. Therefore, another key element in the experimental design was word order manipulation. This section describes the major word order differences between English and Mandarin Chinese, the material selection, and finally the scoring procedure before presenting the data. As the experimental procedure varies according to each specific design, it will be described separately for each experiment.

3.1.2 Word order differences and language interpreting

Word order difference is one of the major topics in contrastive linguistics, but however obvious the difference is across languages, it does not seem to be a popular topic in interpreting studies. Among available studies, attention has been focused on simultaneous interpreting between languages with rigid and flexible word order, e.g., English vs. Japanese (Mizuno, 2005) or English vs. Hebrew (Shlesinger, 2000). These two studies suggested that word order difference between languages may cause an overload of information that was thought to be buffered in an interpreter’s working memory until the head-final verb appears, and only then does the interpreter have sufficient material to construct a TL sentence. The overload, then, could undermine interpretation quality in terms of the completeness of preserving information, because an interpreter's representation of the information held in working memory may decay over time (Mizuno, 2005; Shlesinger, 2000). Shlesinger (2000) manipulated word length (2- vs. 4-syllable adjectives) and the number of the noun modifiers (ranging from one to four) in an SI task. Professional interpreters' English-to-Hebrew interpretation showed a floor effect. The mean proportion of participants' retention for noun modifier was just over 30% when there was only one modifier. When there were three modifiers, their recall dropped to 5% irrespective of the word length of the noun modifiers. Although this result made hypothesis testing difficult, it shows that SI was extremely demanding even for professional interpreters. It also implied a possibility that memory recall for content words like these modifiers was traded for other aspects of the task. One explanation could be that concurrent comprehension and production overloaded these interpreters and resulted in an insufficient capacity for a list of unrelated adjectives. This line of reasoning can extend to a potential competition between grammatical encoding in TL and memory recall in discourse interpreting. Shlesinger (2000) used a serial recall paradigm to address the question as to whether there was a word length effect in SI and whether this effect would be abolished due to interpreters' concurrent SL listening comprehension and TL oral production as a form of articulatory suppression. In contrast, this thesis seeks to explore whether working memory is involved in discourse interpreting by testing whether structural differences (word order rather than word length) between languages affected interpreting performance in terms of discourse content reproduced in the TL. At the end of the last chapter, the hypothesised theoretical rationale for a potential word order effect was delineated and predictions were made regarding participants' performance. The key hypotheses and predictions will be summarised just before the experiment section. The following section briefly discusses major word order differences between Chinese and English, some of which were used as criteria to select experimental discourses.

Word order differences between English and Chinese can be discussed in two dimensions. As far as basic word order is concerned, English and Chinese are Subject-Verb-Object (S-V-O) languages, although whether or not Chinese is a strictly S-V-O language remains a topic of debate[1]. The reason for the assertion in this thesis of Chinese as an S-V-O language is a pragmatic one. The basic word order does not concern this thesis, as experimental sentences are all S-V-O. But when word order differences at phrasal level are considered, the two languages differ dramatically in several aspects (L. Li, 1998). The most salient three categories are: modifiers, associative phrases linked by *de*, and adverbials. These syntactic features were used in selecting discourses for experiments 1-4.

NP Modifiers

In relation to the position of modifiers, both pre-modification and post-modification are allowed in English. However, it is almost always compulsory that modifiers precede modified NP in Chinese.

15. modifier + NP: a(一張) small(小) round(圓) table(桌)
16. NP + modifier: a(一張) table(桌) which is(是) small(小) and(且) round(圓)

As Chinese premodifies an NP, the Chinese translation of phrase 16 would sound very unnatural in daily conversation, although they are allowed and can be found in written material. English NP modifiers can become very complicated when relative clauses (RC) are introduced to modify an NP postnominally. Example sentence 16 is an object relative clause (object RC), in which the modified NP serves as the subject of the matrix clause. Example sentence 18 is a subject relative clause (subject RC), in which the subject, policeman, of the relative clause also serves as the subject of the matrix clause.

[1]Readers are referred to C. N. Li and Thompson (1981) for a succinct introduction as to why deciding the basic word order for Mandarin Chinese is not straightforward.

17. The thief that the policeman chased fell.

警察(policeman) 追趕(chased) 的(relativiser) 小偷(the thief) 跌倒了(fell).

18. The policeman who chased the thief fell.

追趕(chased) 小偷(thief) 的(relativiser) 警察(policeman) 跌倒了(fell).

Associative Phrases linked by *de*

Generally, the word orders of adjective-noun phrases in English and Chinese need not differ, e.g. red(紅) flower(花) or red(紅色) de(的) flower(花), with the genitive marker *de* (abbreviated as GEN hereafter) optional (see C. N. Li & Thompson, 1981, p. 123). But when the first NP modifies the second NP in a noun-de-noun compound, the word orders between English and Chinese can differ. The example phrases are taken from the testing material of Experiments 1-4:

19. hopes for success 成功(success) 的(de) 希望(hopes)
20. the difference of sizes 尺寸(size) 的(de) 差異(difference)
21. importance for the animals 對(for) 動物(animals) 的(de) 重要性(importance)
22. main reason for the decline in the marine species 海洋物種(marine species) 減少(decline) 的(de) 主因(main reason)

Note that there are also many NP-de-NP phrases whose word orders do not have to change, especially when the two noun phrases can form a compound NP. For instance, *China's population* can be translated as 中國(China) 的(de) 人口(population) or 中國(Chinese) 人口(population), and in both cases word order does not change across languages. But the translation of the same phrase from Chinese into English can be either 'the population of China' or 'China's population', with the former showing a different word order. The options that are contingent upon the direction of translation were thought to increase noise in participants' response, therefore these types of phrases were not included in the testing material in the first series of experiments.

Adverbials

"Adverbials are words or expressions that modify verb phrases in the same way that attributives qualify nouns, and they are therefore placed immediately before the verb they modify" (Yip & Rimmington, 2004, p. 138) (also see Sun, 2006 and C. N. Li & Thompson, 1981, Chapter 23). According to Yip and Rimmington (2004), adverbials can be descriptive adverbials and restrictive adverbials. The former describes the manner in which an action is being carried out. Restrictive adverbials on the other hand are used to specify duration, location, tone, structural orientation, or referential scope of verbs. Restrictive adverbials include: 1) time expressions; 2) referential adverbs; 3) set expressions used as mood or tone-setters of an utterance; 4) negators; and 5) coverbal expressions[2].

Again, phrases that are flexible in word orders, i.e. when an adverbial have more choices of position in a sentence in one language than in the other, were not included in the testing material. Also, it was made sure that each sentence in a discourse did not have multiple adverbials. The phrases below are examples taken from the experimental sentences in the first four experiments.

23. duration adverbial 在(zai4)

English: half of them would cease to be spoken within a century

Chinese: 其中一半(half of them) 在(zai4)一個世紀內(within a century) 就沒人說了(would cease to be spoken)

24. location adverbial 在(zai4):

English: annually, 30,000 people in(在) the U.S. take their own lives

[2] It has been claimed that Chinese is the only language that has coverb constructions. Its functional equivalent in other languages has not been decided due to its mixed characters of verb and preposition. "They are called coverbs because they almost invariably have to be used in conjunction with other verbs in a sentence."(Yip & Rimmington, 2004, p. 159) For instance, in example sentence 3, the coverb 爲(wei4) is bundled with an NP 'future generations', followed by the VP of the main clause 'to protect the world's fisheries.' The general formula of S + coverbal expression + main verb hints that coverbs are like prepositions in some cases. But in other cases, coverbs can act like independent verbs: 到(dao4) in the sentence *我們常到(dao4)紐約去*(We often go to New York.)

Chinese: 在(zai4)美國(in the U.S.) 每年(annually) 有三萬人(30000 people) 自行結束生命(take their own lives)

25. coverb 爲(for):

English: implement policies to protect the world's fisheries for(爲) future generations

Chinese: 實施政策(implement policies) 爲(for) 後代子孫(future generations) 保護世界漁業(protect the world's fisheries)

26. coverb 對(towards, to):

English: doctors may want to pay more attention to the relatives

Chinese: 醫生(doctors) 最好(may want to) 對(to) 家屬(the relatives) 多加留意(pay more attention)

27. coverb 比(than):

English: spring flowers bloom a week earlier than they did 50 years ago

Chinese: 春天的花(spring flowers) 比(than) 五十年前(50 years ago)提早了一週(a week earlier) 綻開(bloom)

3.1.3 Material Selection

Eight texts were selected from the Chinese version of the *Scientific American*[3](科學人). Because topic familiarity (Hartley, 1993; Panico & Healey, 2008) and syntactic complexity (Kemper, 1987) are known to affect discourse memory among healthy adults, the material selection was based on criteria of relatively novel topic and simple syntactic structure. It was hoped that these texts would be equally novel to the participants so that participants were less likely to rely on their prior knowledge to reconstruct the text during prose translation. The Chinese version of each text was the translation by professional translators, most of whom were both writers

[3] *Scientific American* is a popular science magazine which publishes 19 editions of different languages around the world. The first issue of traditional Chinese edition was released in 2002.

and scientists. The preparation of the text recording involved a native Mandarin speaker from Taiwan and a native English speaker from the U.K. Half of the texts contained sentences whose word order differed between Chinese and English, and the other half of texts could be translated in such a way that there would be no word order difference between the two languages. The differences of word order in the testing material do not fall particularly into any one category outlined earlier, but the material had more differences as a result of using adverbials. Unlike adverbial constructions using co-verbs, associative phrases and NP modifiers allow more flexible word orders, e.g., 中國的人口(China's population vs. the population of China) and 紅色的花(red flower vs. the flower that is red). The preference of using more adverbial constructions to NP modifiers and associative phrases was meant to make sure that SL and TL versions differ in their word orders.

For both language versions, each text consisted of 14.5 idea units (*Range* 11-18, *SD*= 2.6). For the English version, each text had an average of 100 English words (*Range* 86-117, *SD*=11.1) and the average length of recording was 21 seconds. On average, each Chinese text had 164 characters (*Range* 152-175, SD = 7.5). As it is known that input speech rate can influence prose recall (Wingfield & Stine, 1986), and it has been suggested that input speech rate should not be too fast in interpreting expository discourses (Nida & Taber, 2003), the two speakers were asked to slightly slow down their speech during voice recording for the experimental texts[4]. Four lists of testing material were generated by applying a Latin Square design, so that each participant would be presented with only one language version of each discourse, but each discourse would be translated in both directions of English-Chinese and Chinese-English.

[4]See Appendix for full testing material

3.1.4 Scoring Procedure

As mentioned in Section 2.4.1, the propositional analysis system by A. Turner and Greene's (1977) was chosen as the main analytical tool for these experiments for its clear and detailed instruction in the manual. Another reason that this system is suitable is because it was developed on the basis of Kintsch's (1974) propositional theory. Since Kintsch and Van Dijk's (1978) CI model is one of the main theoretical building blocks of this thesis, a system that is the extension of the background theory should make the theorisation, testing, and interpretation coherent and consistent. Described below are the major principles of discourse propositionalisation with examples selected from experimental discourses. This section concludes with a descrption of how participants' responses were scored.

Discourse Propositionalisation

Recall that a proposition was defined earlier in Section 2.4.1 as composed of a predicate and a number of arguments. The type of relation that a predicate indicates can divide most propositions into three classes: predication, modification and connection.

As predication expresses an action or a state, the predicate in this type of relation is usually a verb and its arguments are taken up by cases that fill the slots of the verb frame. For instance, the phrase *Jane gave Joe a present* can be represented as (GIVE, JANE, PRESENT, JOE), in which GIVE is the predicate, JANE is the agent, PRESENT is an object, and JOE is a goal of state or action identified by the verb. The agent, object, goal, instrument, source, and experiencer are major categories of cases in Fillmore's (1968) case grammar. During propositionalisation, these categories can guide a researcher to identify or categorise the kind of relation he/she finds in a sentence. The other two common predication types are nominal and reference propositions. The phrase *the tongue is the main source of bad breath* has a nominal proposition (IS A, SOURCE, TONGUE). A reference proposition can be used when the relation between cases is not explicitly expressed like a nominal proposition but two cases refer to the same entity. According to A. Turner and Greene (1977), the use of a reference proposition involves inference making, and in fact, A. Turner and Greene (1977) proposed that a text base that serves as a template for research should ideally contain propositions that are directly expressed and also those that are inferred, the justification being that comprehenders can and do make inferences when constructing a coherent representation. This position was not taken in this thesis because it is not clear as to: 1) how many inferences can be made from a discourse; and 2) whether different comprehenders make different inferences. This is discussed later in the paragraph delineating the scoring procedure.

Four types of propositions fall into the category of modification proposition. They are qualifiers, quantifiers, partitives, and negations. Qualifier proposition can be applied to relations between an NP and its modifier, e.g., adjective, or between a VP and an adverb. For example, the phrase *sudden climate change* in the testing material can be represented as (QUALITY OF, CLIMATE CHANGE, SUDDEN). And *climate change has struck the earth many times* can be represented as P1:(STRIKE, CLIMATE CHANGE, EARTH) and P2:(QUALITY, P1, MANY TIMES). In the latter example, P1 is embedded in P2. Quantifier propositions have similar form to that of qualifiers, and they are used to indicate the amount or the extent of a single entity. One quantifier proposition (EXTENT OF, REDUCTION, CONSUMPTION) can be identified in the testing sentence: *The restricted regime generally involves reducing an individual's food consumption by 30 to 40 percent.*

Finally, connective propositions are used to indicate a coherence between propositions. They correspond rather obviously to the adverbials in their surface form. A. Turner and Greene (1977, p. 25) listed eight types of connective propositions:

1. CONJUNCTION: to express union, association, and combination

2. DISJUNCTION: to express alternatives, opposition
3. CAUSALITY: to express cause and effect, correlated events
4. PURPOSE: to express purpose, reason, intent
5. CONCESSION: to express admission of a point, yielding
6. CONTRAST: to express divergence, comparison
7. CONDITION: to express prerequisites, restriction, qualification
8. CIRCUMSTANCE: to express time, location, manner or mode of action

(A. Turner & Greene, 1977, p. 25)

As an example, here is a disjunction proposition (DISJUNCTION: apart from, P1, P2) in a testing sentence: *Apart from the mouth, another common source is the nose,* where P1 and P2 stand for (IS A, SOURCE, MOUTH) and (IS A, SOURCE, NOSE). The whole set of propositions for this sentence is complete with a qualifier proposition (IS, SOURCE, MAIN).

These three categories of propositions were believed sufficient to generate rather consistent text base given a clear instruction and some practice (A. Turner & Greene, 1977). In order to make sure that the experimenter himself was not biased in discourse propositionalisation, twenty sentences were randomly sampled from the whole discourse set and subjected to the propositionalisation by the author and an independent rater who had received training for an hour and had practised on one short passage that was unrelated to the testing material. Inter-rater reliability achieved 91% agreement, and all inconsistent items were discussed and reconciled. The result of the discussion was the list that became the template for response scoring.

Scoring

Participants' oral responses were transcribed and compared against the template textbase. A proposition was fully credited when its predicate and arguments were all present. When either only a predicate or part of the arguments were present, it received half credit. By following D. C. Rubin (1978), synonyms used in any part of a proposition were accepted regardless of tense and agreement (e.g., subject-verb agreement). When a proposition did not correspond to the template unequivocally, its score was decided by the extent the gist of the original proposition was contained in the response (Dixon, Hultsch, Simon, & Von Eye, 1984). Turning to an earlier consideration of the issue of inference making during propositionalisation and scoring, it is acknowledged that there are issues about the finite number of inferences one can make out of a discourse, and about individual differences that result in divergent inferences. Following Dixon et al. (1984), inferences, errors, and elaborations were ignored. Again, the scoring procedure also involved an independent rater in determining whether the experimenter's scoring was biased or consistent. The same procedure in propositionalisation inter-rater reliability test was repeated for response scoring. This time, a transcribed discourse was randomly sampled and scored by the experimenter and an independent rater. The reliability was 89% agreement and the disagreements were resolved by discussion, and the experimenter carried out the rest of the scoring for all responses.

3.2 Experiment 1

The research question that Experiment 1 was designed to address was whether working memory has a role to play in discourse interpreting. The working hypothesis was grounded on the assumption that working memory is required in discourse interpreting in which SL discourse content has to be temporarily stored on the one hand, and on the other hand, a retrieved proposition requires cognitive resource in the TL grammatical encoding. If the single resource account of working memory holds, then participants' limited capacity would be stretched when either storing or processing function is particularly demanding, and therefore the resource competition would lead to decrement in storage or TL production. In the present experiment, the locus of the most resource-demanding operation was assumed to be grammatical encoding where lexical items are sequenced during linearisation (or positional) processing. Two factors were introduced in the design in the hope to cause different resource demand on participants' grammatical encoding. One was the translation direction, and the other, word order differences between Chinese and English. As grammatical encoding has been thought to be non-automatic in a few studies (e.g., V. S. Ferreira & Pashler, 2002), and more so when it is a speaker's L2 (De Bot, 1992; Poulisse, 1997), one way of exploring whether working memory was implicated was to create a situation which may further stretch participants' working memory capacity. The rationale was that the more resource the process function of working memory demands, the less becomes available to its storage function, hence stored information could become irretrievable.

The experimental design capitalised on the potential linearisation problem a speaker can have when retrieved propositions cannot be mapped directly onto the surface structure of a TL, because a direct mapping might change the causality or temporal relation. In the case of interpreting, it was not causal or temporal relation, but the positional relation between retrieved propositions that could cause interpreters' linearity problem. Specifically, it was assumed that resource demand can be higher when a certain surface structure is not shared between languages for a given expression (see Section 2.4.3 for details). Thus sentences that differ in word order between Chinese and English (incongruent sentences) would tax more resource than the sentences that did not differ in word order (congruent sentences). So in a two-by-two design, language and word order were crossed. As regards the prediction, the major effect would be the word order effect, i.e., congruent condition associated with proportionally more propositions preserved in translation than an incongruent condition. It has been found that input language led to poorer discourse memory in bilinguals' L2 than that in their L1. However, grammatical encoding is thought less effortful in L1 than in L2, implying that resource can be largely dedicated to memory maintenance, hence the predicted better performance in memory. The translation directionality issue made predictions difficult for the language effect. Although the directionality effect was hard to predict, the asymmetric resource demand required by grammatical encoding in L1 and L2 promoted a prediction that participants' interpreting performance might be differentially affected by the word order depending on the translation direction, i.e. word order and translation direction might interact. In order to explore the role that working memory played in language interpreting, a measure of memory capacity was taken with Reading Span and Speaking Span tasks. Although both tasks may be highly correlated, they have different foci and demand on the processing component of working memory. Including these measures in the data analysis could inform which measure may be the better predictor of interpreting performance.

3.2.1 Method

Participants

Twenty-four native Mandarin speakers were recruited from the University of Edinburgh for this paid experiment. Each was paid 6.2 pounds. There were two male and twenty-two female postgraduate students, aged between 22 and 30 (M = 26). They all had IELTS[5] score of 6.5 or higher at the time they applied for their postgraduate study (M =7, SD=.56). Although they did not have any training in language interpreting, their IELTS performances suggested that they had an effective command of English. The decision of recruiting Chinese-English late bilinguals is supported by the survey shown in Section 2.3.2 according to which many bilingualism and several interpreting studies had recruited late bilinguals as experiment participants.

[5]IELTS stands for the International English Language Testing System. It is designed to assess the language ability of candidates who need to study or work where English is used as the language of communication. It covers four language skills – listening, reading, writing and speaking.(Council, 2009)

Material

Please see Section 3.1.3 for a description of material preparation.

Procedure

Participants were informed in the recruitment advertisement that this was a study about memory for short stories and the task involved their Chinese and English language skills. The experiment was conducted individually for each participant in a computer laboratory without distractions. The experiment started by showing the instruction:

> This study is about human being's memory for short stories. It is not designed to test your verbal ability but to study how a passage becomes memorised. Please listen to each passage carefully. Immediately after the presentation of each text, please retell the passage as complete as possible in the language other than the presentation language. For instance, when you hear a Chinese passage, you have to retell the passage in English, and vice versa. There will be two practice trials before real testing begins.

In each practice as well as real trial, a passage was auditorilly presented using E-Prime 1.1. During passage presentation, the screen was white, but once the presentation was finished, the screen turned blue. The colour switch was used as a cue to prompt participants to start their TL output as quickly as they could. However, there was no time limit within which they had to finish interpreting each passage. Their responses were recorded in a digital format for later analysis. After the interpreting task, participants were required to complete an Automated Reading Span task (Conway et al., 2005) and an English Speaking Span task.

The Automated Reading Span (RSPAN) was developed by Randall Engle (Engle, 2009) and is available for download on the website of his laboratory in the Georgia Institute of Technology. Its procedure is very similar to the traditional reading span task (Daneman & Carpenter, 1980), but it features computer-paced sentence presentation and randomisation of set sizes. Moreover, memory items were letters, rather than the last word of each sentence to be memorised and recalled. The computer-paced sentence presentation switched the screen from a sentence to letter presentation after a limited period of time. This period was the individual participant's average reading time plus 2.5 standard deviations, calculated during the practice phase. For each set of sentences, participants had to repeat the procedure described below.

Each participant was required to judge whether a sentence was sensible or not as fast as they could by clicking on one of the boxes that were labelled 'True' or 'False' on the screen. Immediately after a judgement was made, a random letter would pop up on the centre of the screen for 800 milliseconds, after which, another sentence appeared on the screen and so on. When the presentation of an entire set of sentences finished, a participant was then required to recall the set of letters by clicking the mouse on his/her memorised letters in the order in which they were presented among a matrix of letters on the screen. Participants were given their judgment accuracy averaged across trials that they had finished right after each response they made, and they were encouraged to keep their judgement accuracy at eighty-five percent correct or above. There were three trials for each set size, and the size ranged from three to seven. Therefore, there were a total of 75 letters and 75 sentences for the judgement task. When a participant successfully completed RSPAN, it generated five values, namely RSPAN score, total number correct, errors, accuracy errors, and speed errors:

- RSPAN: Sum of all perfectly recalled sets
- Total number correct: Total number of letters recalled in the correct position
- Errors: Total number of errors made in judgement.
- Accuracy Errors: Total number of instances where participants verified the sentence incorrectly.

- Speed Errors: Total number of instances where participants ran out of time in verifying a given sentence.

In the Speaking Span task (Daneman & Green, 1986), participants were shown a series of unrelated words on the screen, one word at a time for one second. At the end of the word list presentation, they were required to generate aloud the sentences by using the words in the presented list. Note that each sentence can only contain one item from the list. For example, if the first word was *ignore* followed by *comment* in a set, a participant could say *The teacher ignored the naughty student's bad language* and *The spokesman made no comment on today's breaking news.* They were instructed to use the exact form of each word, and the word could appear in any position in a sentence. They were encouraged to generate sentences in the order in which the words were presented, but if they could not remember the order, sentences could be generated in any order. It was emphasised that participants should not use the last word from a list to start generating sentences, unless that word was the only word they could remember. List set-size increased in an ascending order from two to six, and there were five trials in each set size. Similar to the conventional reading span, two scores could be calculated: one was the maximum set size where a participant could successfully generate sentences for all words from a given list in at least three out of five trials. The other score was the total number of sentences correctly generated using words presented in the entire task.

3.2.2 Results

Before conducting the analysis, one extreme data point was identified, suggesting this participant omitted too much information in discourse interpreting. As it was not possible to determine whether it was due to language proficiency, poor attention or experimental manipulation, his data had to be removed. The dependent variable was the proportion of correctly reproduced propositions in participants'

TL translation. The main independent variables were congruency (congruent vs. incongruent word order between SL and TL) and translation direction (English-Chinese vs. Chinese-English). In addition, four scores were prepared as covariates. Two of them were taken from the Reading Span: total number of recalled letters regardless of their correctness in order, and total number of recalled letters in their correct order. One score was taken from the Speaking Span: total number of words that were used to generate grammatical sentences. And finally, in order to see whether interpreting performance was related to the word order differences, another covariate was taken, called absolute word order difference, measured by the number of places that words were moved from their original to their new position in a TL phrase. For instance, if a phrase consisted of three elements A-B-C in a SL, given that there happen to be TL equivalents that correspond to A, B, and C individually, but TL word order is C-B-A, the word order difference would be 4 places. This is because B's position does not change, but A and C move two places each.

The results first present the mean and standard deviation of participants' interpreting performance in terms of percentage correctly reproduced propositions (Table A.1). A two-by-two repeated measures ANOVA revealed a main effect of word order, $F(1,23)=58.9$, $p<.01$, an effect of translation direction, $F(1,23)=15.2$, $p<.05$, and also an interaction, $F(1,23)=11.23$, $p<.05$. Figure 3.1 shows that memory performance in interpreting was better when translating from participants' L1, Chinese. Participants' performances were also better when the word order differed between English and Chinese. The interaction was driven by a larger congruency effect in Chinese-English than English-Chinese interpreting (see Table A.1 for mean values in Appendix A).

It was not possible to orthogonally cross word order and translation direction for items (discourses that had an incongruent phrase word order cannot be manipulated and became congruent in word order between languages), therefore a traditional by-item analysis could be only used to show whether translation direction had an effect on participants' memory performance. A repeated measure ANOVA showed that the effect of direction was marginal, $F(1,7)= 4.65$, $p=.07$, with a numerical difference of better performance interpreting from participants' L1, Chinese.

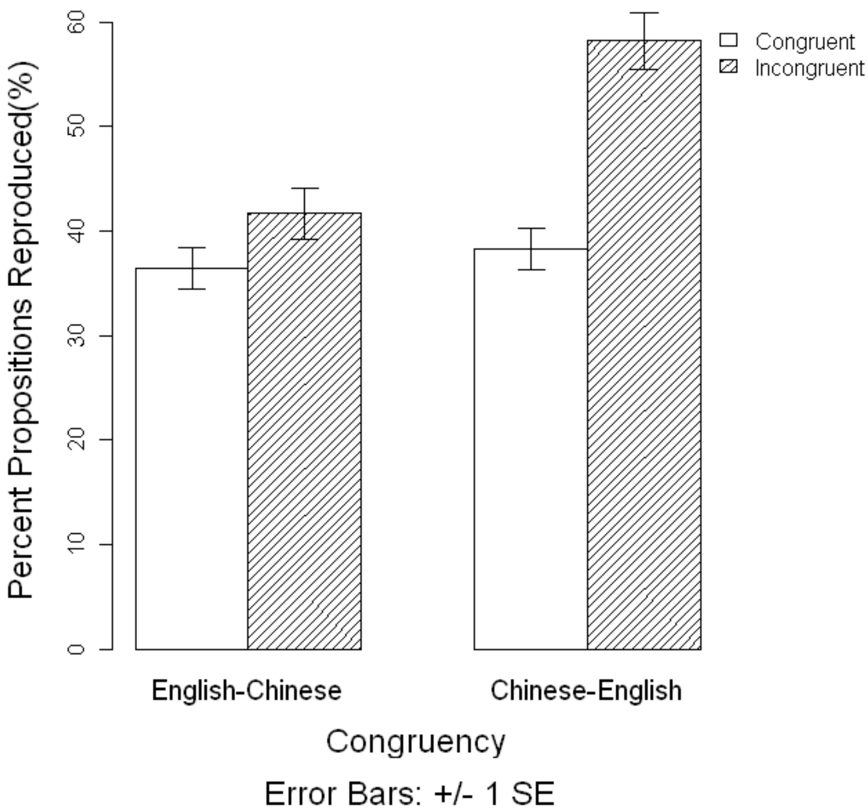


Figure 3.1: Proportion of reproduced propositions as a function of word order congruency.

The second stage of data analysis was primarily exploring whether or not working memory capacity has a role in the task. Several repeated measure ANCOVAs were conducted by adding a different covariate each time. The first covariate added was the working memory capacity measured using Conway et al.'s (2005) automated reading span task: the total number of letters recalled in their correct order in which they were presented. This ANCOVA yielded two main effects. Memory performances were better when word orders were different than when they were not, $F(1,22)=13.35$, $p<.01$, and also when translating from participants' L1, $F(1,22)= 4.69$, $p<.05$. However, the two factors did not interact.

In another ANCOVA, the covariate was participants' working memory capacity measured as the total number of letters recalled irrespective of their order. The effects of word order, $F(1,22)=3.23$, $p=.086$, and translation direction, $F(1,22)=3.49$, $p=.075$, were only marginal. And the initially significant interaction was completely abolished. When participants' Speaking Span measure was added in the equation as a covariate, there were no main effects or interactions.

3.2.3 Discussion

Experiment 1 set out to test whether interpreting performances in terms of correctly reproduced propositions of a discourse was a function of word order difference and translation direction. The predictions made for the experiment were 1) word order difference in discourse between SL and TL might lead to poor memory performance; 2) given that word order did have an impact on the work load of working memory during TL production, the difference in participants' memory performance between congruent sentences and incongruent sentences should be much larger in Chinese-English than in English-Chinese interpreting. This is because the less automatic grammatical encoding in participants' L2 might use a large share of working memory and leave little for the maintenance of discourse content.

Initially, the analysis revealed main effects and an interaction, but the pattern these results showed ran in the opposite direction to the predictions: 1) word order differences were associated with better memory performance; 2) the interaction showed that Chinese-English late bilingual participants were much better in incongruent than in congruent conditions when translating into English, but not when translating into Chinese; and 3) translating into participants' L2 was associated with better memory performance than translating into their L1. Since this thesis considers four components that were thought central to the task, the findings in Experiment 1 are discussed using the same framework and approach.

The locus of the word order effect was assumed to be grammatical encoding, at which stage participants might need to attend to the memory of discourse stored in working memory. The unexpected findings, however, suggested that memory performance in the context of language interpreting was not determined to a great extent by how skillful or automatic participants' grammatical encoding was. Possible determining factors would appear to be other translation components such as comprehension and memory. The process account that led to the prediction that working memory would be shared by concurrent processing in grammatical encoding and memory maintenance seemed to be backed up by the results of ANCOVA where reading span measures and speaking span removed main effects of word order and directionality. It suggested that the capacity measures were mediating factors while the word order difference and translation direction might not have a big role in discourse interpreting. Although the hypotheses were not supported by the data, it gave some directions regarding how discourse interpreting may be explored. One possibility that warranted exploration was that sentences that resulted in word order difference in discourse interpreting somehow involved deeper levels of processing and richer encoding (Craik & Lockhart, 1972), leading to better recall performance. In the levels of processing approach, memory is construed as the record of operations for the purpose of interpreting an array of stimuli. Craik and Lockhart (1972) proposed that deeper analyses of given stimuli would be associated with more durable memory traces than when the analyses were shallower. So it is possible that the effect of word order initially found in the analysis could be due to a qualitatively different parsing strategy used for congruent and incongruent discourses. This possibility was tested in Experiment 3. In the present experiment, only a memory measure was taken. Therefore no evidence was available to indicate that the word order difference increased the participants' work load. Together with the removal of main effects by working memory span, it implied that a revision of hypothesis might have to be made regarding the other element in the hypotheses – memory maintenance. And this has to consider the discussion of memory maintenance and retrieval process in Chapter 2, Section 2.4.2.

The original hypothesis that memory maintenance is resource demanding may be contingent on the question of where and how memory was retrieved. In the review section, an overview of the episodic buffer and LTWM was given as they seem to have rather different proposals for potential mechanisms subserving memory structure and retrieval. The original hypothesis of a role that working memory might play in memory maintenance was more in line with the hypothesis of the episodic buffer, in which discourse memory retrieved from episodic long-term memory has to be buffered in working memory, and the maintenance of this retrieved memory requires cognitive resources. However, the evidence available so far for the episodic buffer (Baddeley et al., 2009) has a focus on the prose verbatim memory, rather than memory at the propositional level. On the other hand, Kintsch and Van Dijk's (1978) CI model and Ericsson and Kintsch's (1995) LTWM construe discourse memory as a retrieval structure that is concomitant with discourse comprehension. Its retrieval is thought to depend on the retrieval cue, as opposed to the propositions themselves, although retrieved propositions can become retrieval cues as well. Discourse memory was conceived as a retrieval structure and retrieval was assumed automatic and effortless (Van Dijk & Kintsch, 1983) unless the structure itself was not well constructed during comprehension or the retrieval cues were not effective. In a sense, memory in the form of a retrieval structure does not seem to require as much maintenance as originally thought. If the retrieval mechanism follows the SAM model (Raaijmakers & Shiffrin, 1981), memory retrieval in discourse recall spreads activation from a retrieval cue in the cue-dependent search process. Since current evidence in psycholinguistics supports incrementality in grammatical encoding, memory retrieval for conceptualisation might also proceed in the piecemeal fashion. The incrementality of conceptualisation and grammatical encoding would imply that a participant does not retrieve everything he/she can remember before starting to speak. To the extent that this reasoning holds, the original assumption that it required participants' resource to maintain the retrieved memory seemed hard to substantiate and it is possible that in language interpreting, most working memory capacity was dedicated to the process components in language production.

However, this revision of hypothesis still cannot explain why the main effects and interaction were removed by the addition of covariates. As ANCOVA can be understood to partial out the variability that cannot be explained by the independent factors, the nature of these covariates would seem pivotal in the hypotheses revision. Linderholm and Broek (2002) found that readers' working memory capacity was correlated more highly, $r(58) = .44$, $p<.01$, with recall, when the the purpose of comprehension was 'study' than when the purpose was 'entertainment', $r(58)=.12$, $p<.01$. If higher capacity of working memory is reliably correlated with better recall, it is consistent with the prediction the CI and LTWM would make for discourse memory. Given that the reason underlying the removal of main effect and interaction by working memory capacity in this experiment is along the line argued in the CI and LTWM model, there is a strong reason to encourage a hypothesis that memory performance in discourse interpreting in this experiment may be determined to a greater extent by the comprehension component.

In light of the unexpected results of Experiment 1, three more experiments were conducted. The first follow-up study was designed to explore the interaction between word order and translation direction observed in Experiment 1. The participants of Experiment 1 were postgraduate students who majored in subjects other than interpreting or translation, therefore, it was likely that their language proficiency played a prominent role in their performance. Participants' much better performance in incongruent than in congruent sentences when translating into English was the most puzzling result of all, because usually unbalanced bilinguals have more receptive than productive vocabulary in their lexicons, especially in their L2. Therefore it might have been easier for them to comprehend than to produce L2 discourses. So the third branch study was conducted among a group of postgraduate students who were trained as interpreters. Given that these students were more advanced in English when they enrolled for the programme, it was thought that they might perform differently from the participants in Experiment 1. The research question of this study was whether advanced proficiency in L2 would remove the effect of translation direction.

Another branch of study set out to address the research question of Experiment 1 from a different perspective. Even though participants' response transcript in Experiment 1 showed more word order differences in translating incongruent sentences, and memory performance was better than in translating incongruent sentences, it was not clear whether the better recall was related to any kind of transformation, e.g., word order change. If translating incongruent sentences does involve some kind of transformation, which enriches the encoding process and leads to better recall, then there is one task that can be used to test this hypothesis: paraphrasing. So the question can be formalised: if the observed word order effect was due to some kind of transformation, it is possible that participants' memory might be better in paraphrasing than in repeating a discourse.

Recall that in Section 2.4.1 where it was discussed how comprehension purpose and input language can affect discourse comprehension. However, in light of the unexpected results of Experiment 1, a factor that appeared in Section 2.4.2 warranted special attention, and that was the syntactic characteristics in discourses. McKoon et al. (1993) proposed that syntactic features might be used by comprehenders to identify the most prominent and crucial propositions and these propositions are more likely to be held in memory and more accessible when retrieval is required. Consistent with the account of levels of processing (Craik & Lockhart, 1972), it was possible that the discourse sentences which differ in word order between languages have the kind of syntactic features to which participants were sensitive, and these sentences were encoded in a more coherent or rich way than the congruent sentences. In this case, maybe sentences that are incongruent in word order between languages were intrinsically more memorable than congruent sentences. This possibility was tested in the third study that followed on from Experiment 1. The three branch studies will be described in detail in the following sections.

3.3 Experiment 2 – Evidence from interpreting students

Experiment 2 was designed to further explore the translation direction effect. These participants were more advanced in English, therefore, it was predicted that they would show no, or a smaller, effect of translation direction than participants in Experiment 1. Even if they showed an effect of word order, the effect would remain constant irrespective of translation direction. So translation direction would not interact with word order in terms of memory performance.

3.3.1 Method

Participants

Twenty-four interpreting students who were Chinese-English bilinguals who learned English as a second language took part in Experiment 2. Sixteen were from the MA Interpreting and Translating course in the University of Bath, and eight were from the same course in the University of Newcastle. There was only one male student. They were aged between 22 and 31 (M=25). Each participant was paid

10 pounds. They had IELTS scores between 6.5 and 8.5 (M= 7.3, SD= 0.5). None of the participants in the present experiment had participated in Experiment 1.

Material

The discourses were the same as those used in Experiment 1. Please see Section 3.1.3 for details.

Procedure

Experiment 2 extended Experiment 1's procedure, but modified the working memory tasks. As the main procedure remained the same, it is not repeated here, but the modification of working memory tasks is described.

There were two main reasons for the task modification. One was related to a technicality of Conway et al.'s (2005) automated reading span. While administering the task, the programme sometimes generated very odd scores. For instance, a participant received a score of 5 for the total number of correctly recalled letters irrespective of their order. This dramatically low score did not seem plausible. As it was not clear how to retrieve the raw score of this participant, it was not possible to confirm the cause of the poor performance. Another issue was that it was a lengthy task, and it usually took about half an hour to forty minutes to complete. Considering that participants have other tasks to carry out, it was decided not to include Conway et al.'s (2005) reading span in the present experiment and to switch back to the conventional reading span task.

Participants always did the interpreting task before memory tasks. The two memory tasks they did were digit span and reading span. Instead of using E-Prime to collect data, a web-based test on a BBC website 'Explore your memory.'(Logie, 2009) was used. This is an ongoing project developed to study human memory with a cross-sectional approach (Logie & Maylor, 2009). The entire web-based test included ten parts, among which were digit span and reading span based on Baddeley, Logie, Nimmo-Smith, and Brereton (1985) and Duff and Logie (2001). In the digit span task, participants were presented with a digit list on the computer screen, one digit at a time, for one second. Once a list of digits was all presented, participants were prompted to recall the digits by typing them in the order they appeared in a blank box at the middle of the computer screen. The list length increased in the ascending order from 3 to 9. There were two trials for each list length. When participants succeeded in recalling either trial in a list length, the test continued until they failed to recall any digit list correctly. The span measure would be the largest list length they could correctly recall the digits for in at least one out of two trials for a list length.

In the reading span task, each trial contained a set of sentences that were presented on the screen, one sentence at a time. Participants had to read and judge each sentence regarding their plausibility as quickly as possible, and then memorise the last word of each sentence. When a set of sentences were all presented, participants recalled the last words of each set of sentences by finding them out from a matrix of words and then draging each word into a box in the original order in which they had appeared. The set size began with 2 and the largest set was 6. Each set size had two trials and the test terminated when a participant failed to recall words correctly for any trial in a set size. The reading span measure in this task was the largest set size that one can correctly recall for at least one out of two trials.

Participants' interpretation responses were transcribed and scored using the same method described in the general procedure in Section 3.1.1, paragraph **Scoring Procedure**.

3.3.2 Results

Two extreme values (Recall=0) were removed from the data set. The independent variables were word order congruency (congruent vs. incongruent between Chinese and English) and translation direction. The dependent variable was the percentage of propositions correctly reproduced in a TL discourse. The two covariates, reading span and digit span, were taken from the web-based memory tests introduced in the procedure section.

Figure 3.2 shows interpreting students' performance in terms of proportions of correctly reproduced propositions (see Table A.2 for mean values in Appendix A). This graph suggested that 1) there was an effect of translation directionality; 2) the two factors interacted as the performances in translating into Chinese in congruent and incongruent conditions were effectively the same; and 3) the variability was smaller in Chinese-English condition than that in the reverse direction.

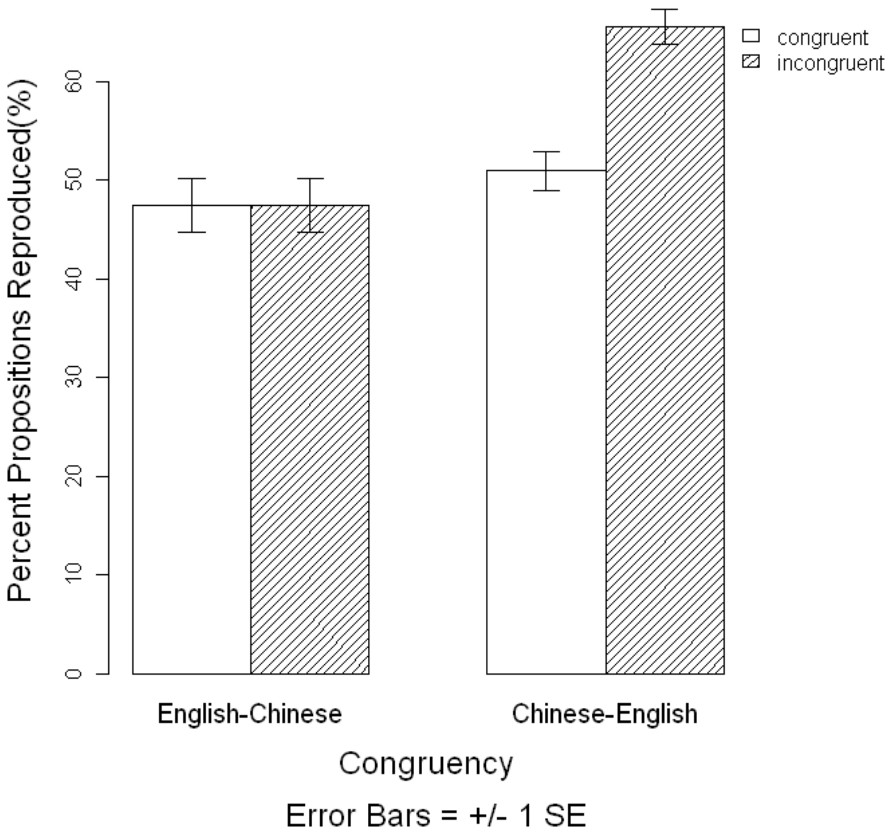


Figure 3.2: Proportion of reproduced propositions as a function of word order congruency.

The data set was subjected to a repeated measures ANOVA in which the repeated factors were word order congruency and translation direction. The test confirmed the effect of translation direction, $F(1,23)=23.4$, $p<.01$, an effect of word order congruency, $F(1,23)=12.1$, $p<.05$, and the two factors interacted, $F(1,23)=24.8$, $p<.01$. This replicated the result of Experiment 1. Participants (in this case, interpreting students) were better when translating from L1 (Chinese) than from English. They were also better when word orders in sentences of a discourse were different between two languages. The interaction was driven by the difference between congruent and incongruent condition within Chinese-English interpreting (Figure 3.2). The by-item analysis revealed that translation directions did not affect memory performance, $F(1,7)=5.25$, $p=.06$.

Two ANCOVAs were conducted in the second stage of analysis. The participants' digit span was added as a covariate in the modelling. As shown in the Table 3.1, digit span abolished the main effects and the interaction observed in the main analysis.

Table 3.1: An ANCOVA using participants' digit span as covariate. The covariate removed the main effects and an interaction originally observed in the initial analysis.

source	df	Mean Sq	F value	Sig
dir	1	108.615	1.501	.233
dir*DigSpan	1	51.982	.719	.406
Error(dir)	22	72.334		
congruence	1	513.282	2.488	.129
congruence*DigSpan	1	350.157	1.697	0.206
Error(congruence)	22	206.283		
dir * congruence	1	70.762	.699	.412
dir * congruence * DigSpan	1	155.052	1.532	.229
Error (dir*congruence)	22	101.198		

In the second test, participants' reading span was used as the covariate. The result showed that the translation direction effect was significant, but not the word order or the interaction between the two factors (Table 3.2).

Table 3.2: An ANCOVA using participants' reading span as the covariate. The covariate removed the main effect of word order and an interaction originally observed in the initial analysis.

source	df	Mean Sq	F value	Sig
dir	1	325.667	4.762	.04
dir*RSpan	1	138.916	2,031	.168
Error(dir)	22	68.393		
congruence	1	69.086	.311	.583
congruence*RSpan	1	.154	.001	0.979
Error(congruence)	22	222.192		
dir * congruence	1	184.474	1.728	.202
dir * congruence * RSpan	1	32.829	.308	.585
Error (dir*congruence)	106.754			

These ANCOVA statistics suggested that there was a considerable share of the variance that was not explained by the independent variables and it appeared that the capacity measure had a role in participants' memory performance.

The last stage of analysis was a contrast of performance between general postgraduate students in Experiment 1 and interpreting students. A mixed repeated measures ANOVA was conducted with group as between-subject variable and word order as well as translation direction as repeated measures. As working memory was measured using different tasks across the two experiments, it was not possible to compare between-group difference on working memory measures. The ANOVA showed an effect of group: interpreting students (M=52.4) were significantly better in recall of idea units than postgraduate students (M=43.5). There was a main effect of word order congruency, $F(1,46)=55.9$, $p<.01$, an effect of translation direction, $F(1,46)=38.3$, $p<.01$, and an interaction between two factors, $F(1,46)=31.1$, $p<.01$. All but the factor of word order interacted with group, $F(1,46)= 4.1$, $p =.49$, indicating that the performance difference between congruent and incongruent conditions was significantly larger in interpreting students than general postgraduate students.

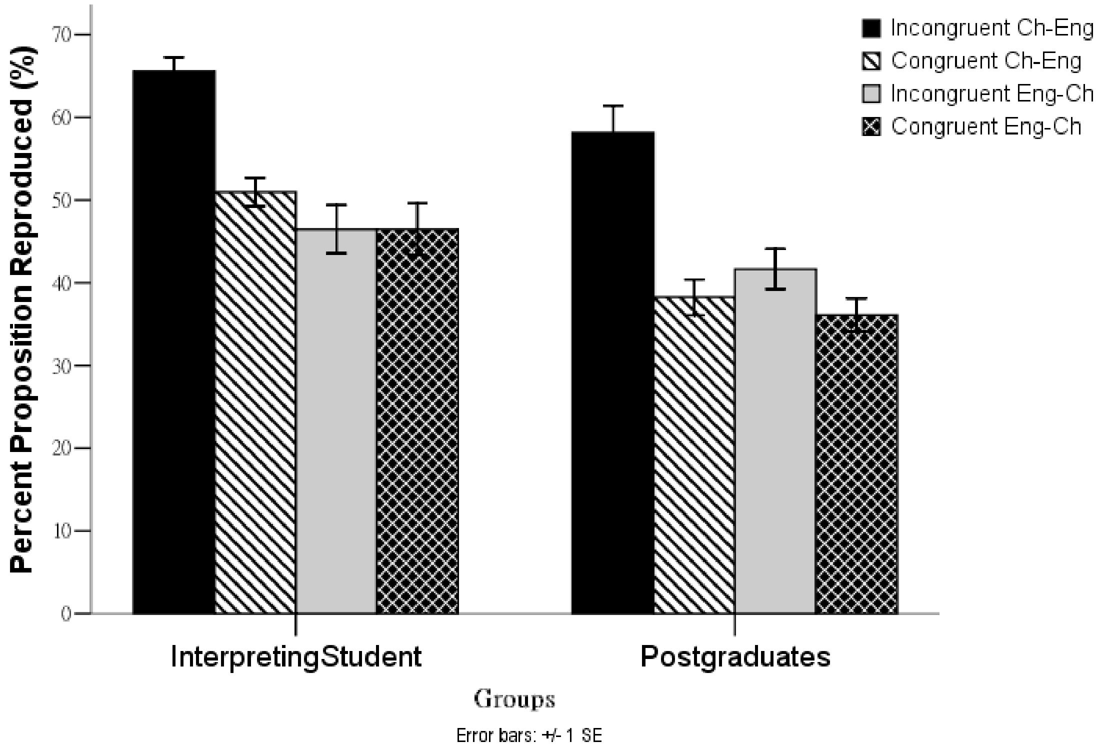


Figure 3.3: Proportion reproduced propositions as a function of word order congruency and group.

3.3.3 Discussion

Experiment 2 set out to address the question of whether participants who were more proficient in their L2 would be less affected by input language and word order. Initially, interpreting students who took part in Experiment 2 showed exactly the same pattern in their data: main effects of word order congruency and translation direction; and the two factors interacted. The conclusion, however, was not so straightforward because the word order effect was removed when the variance accounted for by individual differences in digit span and reading span was partialled out by ANCOVA, and so was the interaction. It therefore appeared that word order and translation direction were important, but it might be their memory capacity that was the determining factor in the task of discourse interpreting.

The contrast between two groups showed a trend that interpreting students' performance was superior to that of postgraduate students. These interpreting students might have been more proficient than postgraduate students before they enrolled for the study in the U.K. as most of them had their first degrees in subjects related to languages. Many of them have been working as freelance translators before their postgraduate training in the U.K. Therefore, they would have been much more familiar with translation as a task than the participants in Experiment 1. In addition, their programmes in interpreting involved intensive skill development, exercise and building a repertoire of strategies in the face of certain constructions. So the overall better performance should come without surprise. In Figure 3.3, it can be seen that neither interpreting students nor postgraduate students were affected by word order difference when they translated discourses into Chinese. But when translating into English, the word order effect appears significant. As mentioned earlier, one candidate account for the advantage of recall of idea units for incongruent discourses could be the levels of processing framework (Craik & Lockhart, 1972). The discourses that resulted in different word order in interpreting between Chinese and English might have certain characteristics that drove participants to process them in a qualitatively different way, perhaps in a semantically deeper fashion, and resulted in durable memory traces. This account is in fact consistent with Kintsch and Van Dijk's (1978) CI model as well. Kintsch and Van Dijk (1978) proposed that discourse memory is incidental to the construction of a coherent mental representation of a discourse. The discourse memory is a by-product of comprehension. There could be prominent syntactical features in the discourses that had incongruent sentences. It was possible that these features were identified by the participants in Experiment 1 and 2. Following McKoon et al. (1993) and Kintsch and Van Dijk (1978), prominent propositions are likely to be selected and placed in working memory in order to construct a coherent representation by linking incoming propositions with those already processed. These prominent propositions themselves were salient and might have gone through cyclical processing according to Kintsch and Van Dijk (1978), therefore, they became more accessible and more easily retrievable when recall is required.

But the question one could ask is why this advantage was not present in interpreting from interpreters' L2, English? This suggests that this kind of advantage might be sensitive to language proficiency. Perhaps the English proficiency of the participants in both experiments was not good enough to take advantage of the syntactic features. Alternatively, as Clahsen and Felser (1996) pointed out, L2 learners typically perform shallow parses only, leading to a kind of encoding not as rich as they would when parsing discourses in their L1.

The fact remains that, when participants' working memory capacity was taken into consideration, the statistics indicated that perhaps word order and translation direction were not determining factors for their memory performance. It is possible that participants with larger working memory capacity can encode the incongruent passages more effectively (deeper level of processing), and therefore, they show a bigger advantage for word order change than do low working memory capacity people.

Nevertheless, this result should not be taken to conclude that word order and translation direction did not influence participants' processing in language interpreting. Perhaps the key processes underlying language interpreting can be better reflected upon by monitoring other performances in addition to memory recall. This will be further discussed in the second series of experiments.

3.4 Experiment 3 – Evidence from discourse paraphrasing

The results of Experiment 1 yielded several possibilities to explore, one of which was whether better memory performance was associated with on-line reformulation of linguistic elements in sentences. This question is different from the fundamental one that concerns whether language interpreting involves manipulation of constituents in real-time. The latter is more relevant to simultaneous interpreting because the temporal proximity between the perception of a SL text and the production of its TL output creates a situation in which on-line 'transformation' or 'reformulation' is more likely to occur than in other situations, such as consecutive interpreting. As there is no evidence available that suggests interpreting involves real-time word order manipulation, it would be inappropriate to assume a causal relation between better memory performance for sentences that differed in word order between languages than those that did not. But in order to find out why incongruent sentences were associated with better memory performances without relying on the assumption that word order difference between languages must imply word order manipulation on-line, the participants were required in this experiment to actively change word form or word order when they recalled a discourse. The task in focus is paraphrasing.

Paraphrasing has been commonly used in SI research to tease apart whether language interpreting is a task with subcomponents of language switch and linguistic reformulation (Christoffels & De Groot, 2004). Malakoff and Hakuta (1991, p. 151) mentioned "translation has been called 'interlanguage paraphrase' or 'intralanguage translation'. In both, the objective is to take a piece of information and recode the meaning in a different linguistic form – in one case the form is a different language and in the other, the form is within the same language". Christoffels and De Groot (2004) compared recall performance in their experiment that crossed tasks (shadowing, SI, paraphrasing) and recall condition (immediate and delayed). The memory in the immediate conditions was the content produced at the time participants carried out their tasks, whereas the memory in the delayed conditions was the recall in a cued-recall task after the primary tasks were completed. In both conditions, recall was measured using a rating system whose scores ranged from 0 to 3. Overall, recall was better in the delayed than in the immediate condition. What is relevant to the present experiment is the result of the delayed condition. Christoffels and De Groot (2004) showed that recall was best in the interpreting condition, followed by paraphrasing, and shadowing. The differences between each condition were significant. Better production quality in paraphrasing than in shadowing has also been reported earlier by Anderson (1994). Christoffels and De Groot (2004) referred to the levels-of-processing (Craik & Lockhart, 1972) account in explaining better memory after paraphrasing than after shadowing. They contended that shadowing may involve the processing at a shallower semantic level while paraphrasing involves deeper semantic analysis of the input. If the same account holds between paraphrasing and free recall, one would expect to find a similar pattern in the present experiment.

3.4.1 Method

Participants

Twenty-four postgraduate students (6 male and 18 female) who were native Chinese speakers were recruited in the University of Edinburgh in this experiment. They were aged between 22 and 30 (M=24.7). None of them had taken part in Experiment 1 or 2. Each participant was paid 3 pounds for this thirty-minute experiment.

Materials

The material used in Experiment 1 was used in this experiment. But the word order congruency here bears no reference to the word order difference across languages – it only serves to categorise these Chinese discourses into two types. Incongruent type of discourse contained several NP or VP modifiers, whereas the congruent type did not.

Procedure

The difference between the present experiment and Experiment 1 was that the language of the material and the language for recall were the same (Chinese). In the paraphrase block, participants paraphrased the presented discourses and in the verbatim block, they were asked to recall the discourses word-for-word. Discourse type and recall method were counterbalanced so that each discourse was used in both verbatim and paraphrase block. The reason that congruent and incongruent discourses were counterbalanced in stimulus presentation was because it was not clear whether one type of discourse was more prone or easier to be paraphrased than the other.

There was one practice trial for each condition before participants started real trials. Discourses were auditorily presented through a headset. During discourse presentation, the computer screen remained blank. Once each discourse presentation was completed, the computer screen turned grey and at the same time, a chime 'DING' was set off through the headset to prompt participants' recall or paraphrase. In the paraphrase block, participants were instructed to change the word order as much as possible without changing the meaning of the discourse or at least to change word forms by replacing words with synonyms or phrases. Participants' responses were recorded digitally for later analysis.

The scoring procedure repeated the one that was described in the general methodology section. As the focus was the memory performance and comprehension purposes, participants' working memory measures were not taken. In order to make data interpretation easier to follow, the term 'recall method' will be used instead of the 'comprehension purpose' or 'reading purpose' when comparing memory performance in verbatim and in paraphrase conditions. 'Recall method' will also be used in the remainder of the thesis when a comparison is made on memory performance between conditions that required participants to recall discourses/sentences with different methods, e.g., repetition vs. translation. Participants' interpretation responses were transcribed and scored using the same procedure described in the general procedure in Section 3.1.1, paragraph **Scoring Procedure**.

3.4.2 Results

In the first stage of analysis, data sets were analysed regardless of whether participants complied with the instructions or not, i.e. assuming that the instructions of paraphrase were followed.

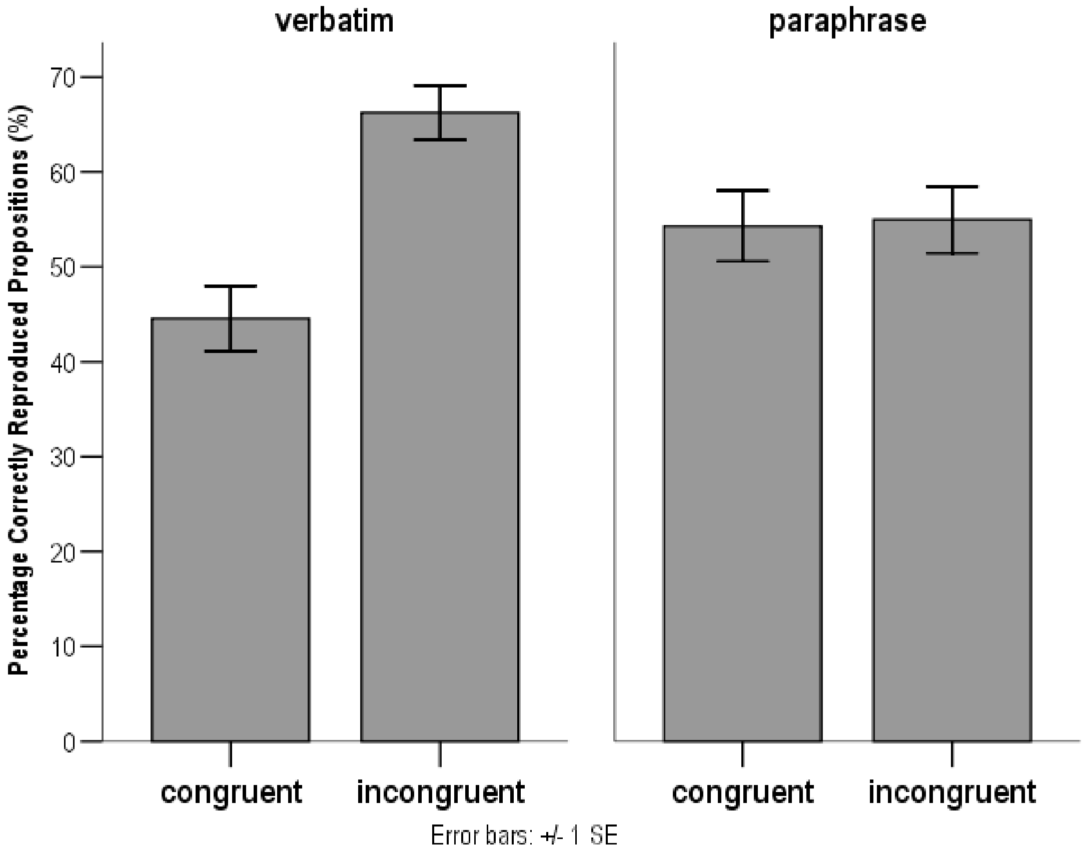


Figure 3.4: Proportion of reproduced propositions as a function of comprehension purpose and discourse type.

Results present the first-stage by-subject analysis which assumes that all participants followed the instructions to paraphrase. The recall method (verbatim recall vs. paraphrase) was the independent variable, and discourse type (congruent vs. incongruent, note the difference does not refer to cross-language word order difference) was the between-subject factor in a mixed ANOVA. Overall, there was no effect of recall method, $F(1,23)=.09$, $p=.77$. However the interaction between recall method and discourse type reached significance, $F(1,22)=16.9$, $p<.01$. Figure 3.4

illustrates that mean percentage recall did not differ between verbatim and paraphrase conditions (see Table A.3 for values in Appendix A). The most interesting result was that the incongruent discourses were associated with better memory performance than congruent discourses were in verbatim recall, but this difference was removed when both types of discourses were paraphrased by participants. In by-item analysis, there was an effect of word order congruency, $F(1,7)= 8.27$, $p<.05$, but neither the effect of comprehension purpose nor the interaction between the two factors was significant. The effect of word order in by-item analysis suggested that incongruent discourses might be somehow more memorable for reasons other than the word order difference between Chinese and English.

Since there was no trend of instruction effect in the first-stage analysis, the raw data from participants was examined and it was noticed that on multiple occasions, participants produced verbatim recall in the paraphrasing condition. So it was decided to collapse and regroup the data according to actual word order changes in each discourse in the second stage analysis.

The data were collapsed across conditions and were grouped by whether or not a sentence was actually paraphrased by examining participants' responses trial-by-trial. A trial was coded as paraphrased if it showed word order change or form change. When there were more changes in one aspect than the other, this trial was labelled according to the dominant type of paraphrase. The paraphrased group of trials were further broken down into two types of paraphrasing: word form vs. word order. This regrouping resulted in three subsets of data: discourses that did not show any type of paraphrasing, discourses that had primarily word order changes, and lastly discourses that had primarily word form changes. Thirty-five out of ninety-six discourses in verbatim block showed form changes (15 discourses) and order changes (10 discourses). Forty-eight out of ninety-six discourses in paraphrase block were paraphrased with form changed (20 discourses) or order changed (29 discourses). This figure suggested that only half of the total discourses in the paraphrase block were actually paraphrased.

This dataset was modelled using univariate ANOVA with several different settings, because trials were not independent of each other in this repeated measure design.

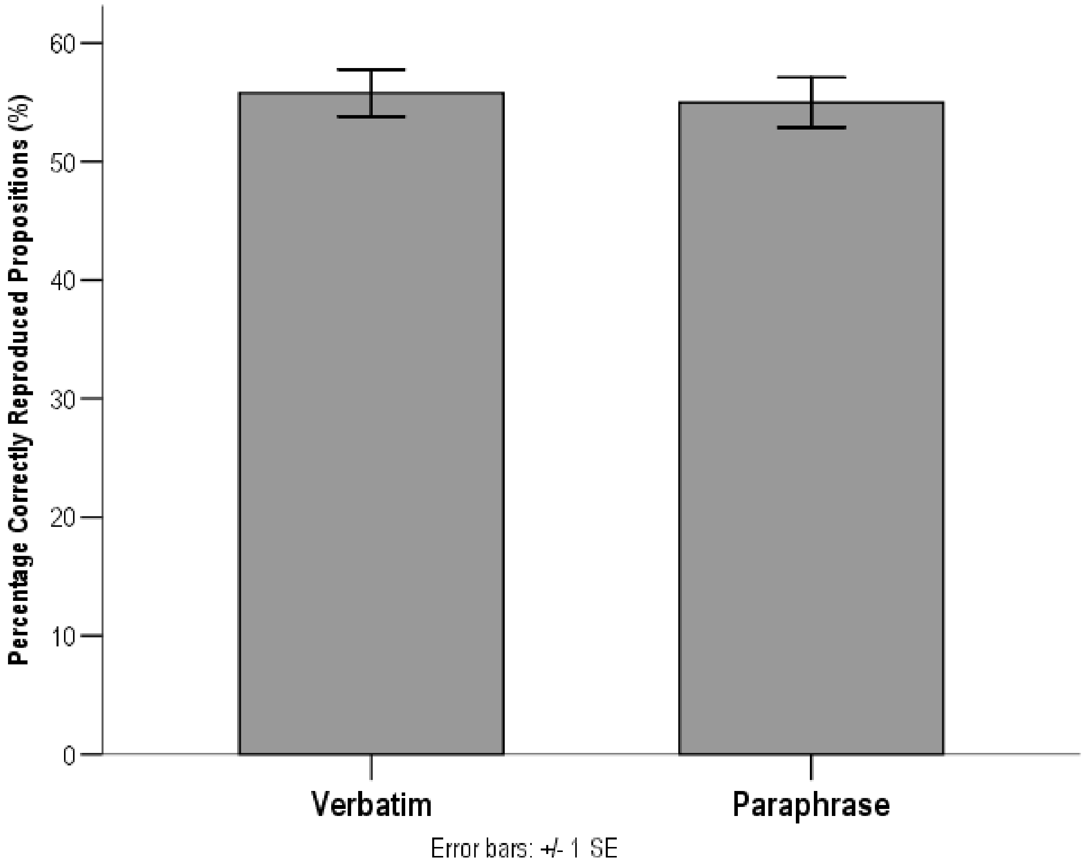


Figure 3.5: Proportion reproduced propositions as a function of comprehension purpose across trials.

In the first stage of modelling, the two types of paraphrase (form change and word order change) were collapsed into one type of response and then contrasted with verbatim responses. The analysis with subject as the random factor showed that the effect of paraphrase did not reach significance, $F(1,23)= 1.4$, $p=.25$. The same result was obtained when discourse was used as the random factor, $F(1,7)= .01$, $p=.91$ (Mean data are shown in Figure 3.5).

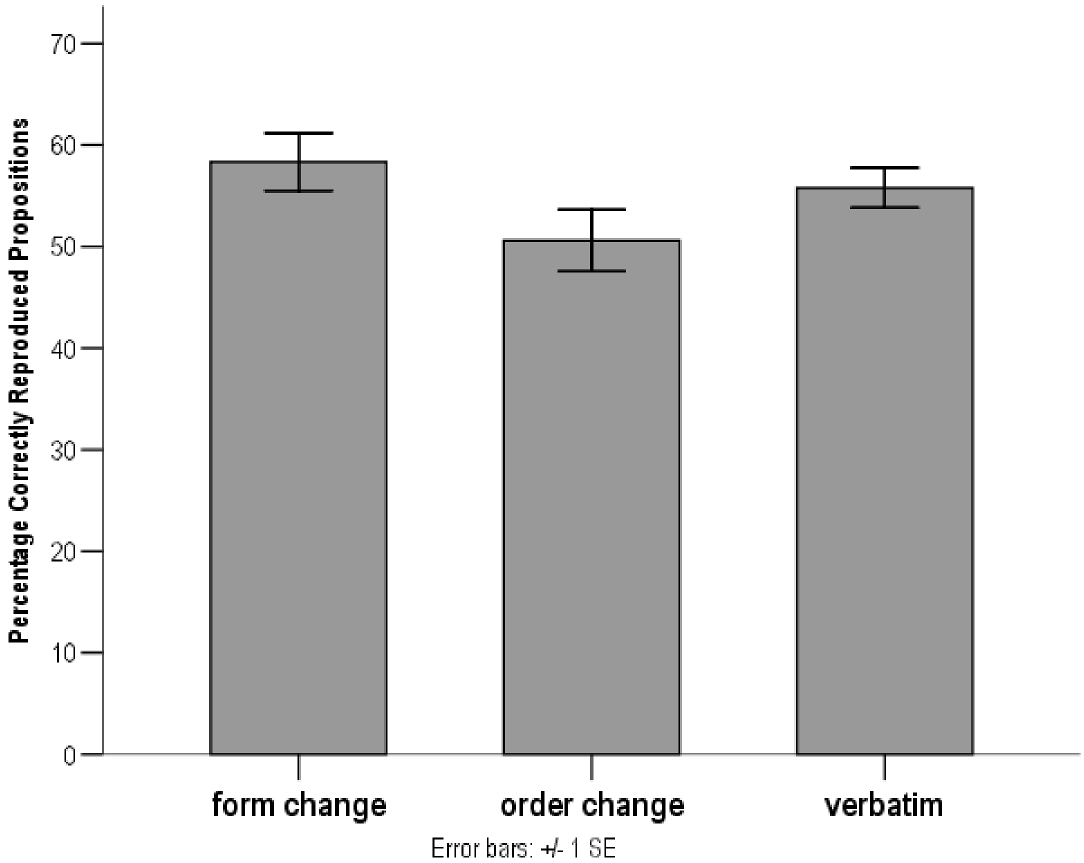


Figure 3.6: Proportion of reproduced propositions as a function of comprehension purpose across trials.

When analysed with the factor of paraphrase having three levels: form change, order change, and no change, again the effect of paraphrase did not reach significance when the model used subject, $F(2{,}44)= 2.6$, p=.09, or discourse, $F(2,\ 17)= 1.1$, p=.37, as the random factor (See Figure 3.6 for mean percentage of recall.).

3.4.3 Discussion

The question of whether active transformation was associated with the better memory performance in Experiment 1 in the incongruent condition was addressed in the present experiment by comparing participants' discourse memory performance in verbatim and paraphrase conditions. It was predicted that if Christoffels and De Groot's (2004) observation that delayed recall after paraphrasing was better than shadowing can generalise to the comparison between paraphrase and verbatim recall, maybe to paraphrase a discourse would lead to better memory performance than to recall it verbatim. This prediction was not supported by the data. Participants' memory performance did not differ between verbatim recall and paraphrase. Had participants in the Experiment 1 carried out some kind of reformulation that involved processes similar to paraphrasing, one would expect to observe compatible performance with the pattern on the right panel of Figure 3.4 regardless of the word order congruency. Given that participants' performances were more similar to the pattern on the left panel, together with the statistical test results that found no relation between structural change and memory performance within a language, it is probable that active transformation that involved word order or form change was not directly related to participants' better memory performance in discourse interpreting.

There have been discussions surrounding the topic of comparing SI and paraphrasing on the grounds that they both seem to require linguistic reformulation in one way or another (Christoffels & De Groot, 2004). Although the study by Christoffels and De Groot (2004) was motivated by identifying what component of SI made it so demanding, their evidence of longer speech initiation for paraphrasing than for interpreting challenged the view that paraphrasing is a within-language version of language interpreting. One key task requirement of paraphrasing is that participants should avoid repeating the original wording. In order to achieve this, they must not only inhibit repetition, the premise for them to be able to do that is an ability to reten the original form or word order of a sentence This is possible when the testing material is a short sentence, as word order priming has been shown to be operative in sentence production (Hartsuiker & Westenberg, 2000). But memorising a whole discourse verbatim is a greater challenge. And this was shown in the data where participants were only able to meet the demand of task instructions of paraphrasing for a small proportion of each discourse. In other words, participants had a great difficulty in actually generating paraphrased recall, even within their first language. Recall data alone is not sufficient to suggest that these changes of word form or word order were actually made consciously, or indeed if they were made at the stage of language production, therefore future research could replace discourse with sentences in addressing these issues.

The findings that incongruent discourses were recalled better in verbatim block in by-subject and also by-item analyses suggested that incongruent discourses might be more memorable than congruent discourses. A separate analysis on participants' actual responses irrespective of the instruction of recall methods established that 1) recall discourse by paraphrasing was difficult and 2) paraphrasing did not lead to better recall than verbatim recall did. This evidence implied that it was possible that participants' better memory performances in Experiment 1 and 2 were not associated with some kind of syntactic or lexical transformation. It appears that the analyses in the first three experiments still left the question open as to whether incongruent discourses were intrinsically more memorable than congruent discourses.

3.5 Experiment 4 – Evidence from free recall

The third branch study based Experiment 1 was motivated by the finding that the effect of word order and translation direction were removed by the factor of memory capacity added as covariate in the data modelling. One key component of translation is comprehension, which plays a central role in the CI model (Kintsch & Van Dijk, 1978) and sentence memory superiority explained by LTWM (Ericsson & Kintsch, 1995). With evidence in Experiment 2 and 3 converging to suggest a possibility that word order might not be a key predictor for memory performance in discourse interpreting, Experiment 4 was a control study to test the hypothesis whether some discourses were intrinsically more memorisable than others.

3.5.1 Method

Participants

Forty native English speakers (8 males and 32 females) were recruited in the University of Edinburgh. It was clearly specified on the recuitment statement that only participants whose first language is English were wanted. There was no questionnaire-taking upon their arrival to check their linguistic background. Each was paid 3 pounds for this 30-minute experiment. They were aged between 18 and 27 (M= 21.6).

Material

Testing materials were the same as those used in Experiment 1-3. Please see Section 3.1.3 for details.

Procedure

The procedure remained largely the same as that in Experiment 1 except that the task after auditory passage presentation was a simple verbatim recall of each discourse in the same language as used in presentation and in the first language of the participants (English). Half of the participants began with congruent discourses and the other half began with incongruent ones. They were not allowed to make notes. Participants' discourse recall was digitally recorded for transcribing and scoring. Participants' interpretation responses were transcribed and scored using the same procedure described in the general procedure in Section 3.1.1, paragraph **Scoring Procedure**.

3.5.2 Results

Figure 3.7 shows that memory (English-English) for discourse with sentences that have incongruent word order between English and Chinese (M=62.1, SD=6.6) was better than that for discourse with sentences with congruent word order between

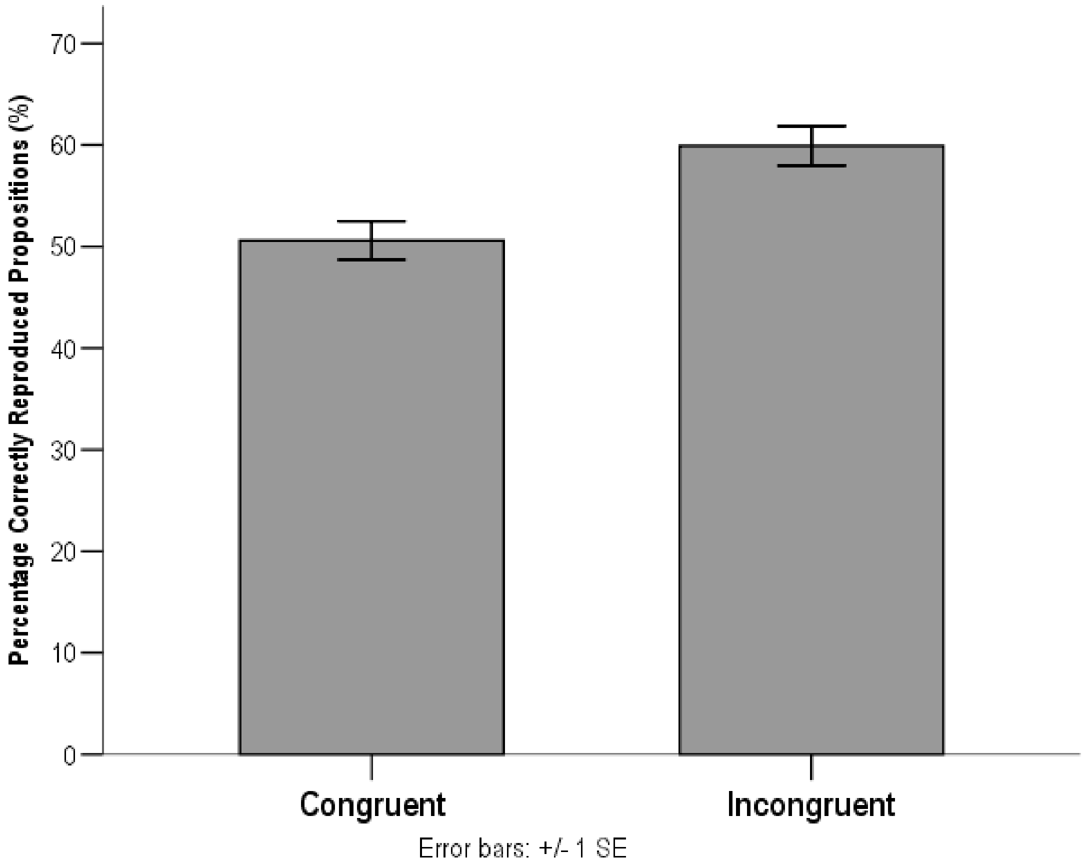


Figure 3.7: Proportion of reproduced propositions as a function of word order congruency.

the two languages (M=50.6, SD=8.6). A mixed ANOVA was conducted using the order of discourse counterbalanced as between-subject variable. In by-subject analysis, there was an effect of word order congruency, $F(1,38)= 35.8$, $p<.01$, but word order did not interact with counterbalanced group. A similar finding was observed in by-item analysis, in which the order effect reached significance, $F(1, 13)= 9.1$, $p<.05$.

Although it has been established in Experiment 3 that active word order or word form change were not significantly associated with memory performance, this was explored further by sampling the ten best trials that resulted in the best memory performance in that experiment and ten most poorly recalled discourses to scrutinise the actual change in wording that participants generated. The ten best recalled discourses did not show a high frequency of word form or order change. However, the content of the recall of the ten most poorly recalled texts deviated considerably from those that participants had been attempting to reproduce.

3.5.3 Discussion

The better memory performance for incongruent sentences than congruent sentences supported the hypothesis made in the discussion of Experiment 1 that memory performance might be determined by the processes pertinent to the comprehension component to a great extent in a free recall task. The implication of this finding and observations across the four experiments will be considered in the general discussion of Experiments 1-4, which follows.

3.6 General Discussion

This series of experiments began with the question, whether working memory has a role to play in discourse interpreting. A full consideration of four translation components identified a potential stage during interpreting where working memory might be required. That stage was assumed to be grammatical encoding. By assuming a unified working memory resource, interpreters' limited capacity would have to be shared by concurrent tasks during grammatical encoding. Grammatical encoding in L2 has been assumed non-automatic (De Bot, 1992; Poulisse, 1997), therefore, the more demanding encoding becomes, the less resource is available to other ongoing tasks. One of the ongoing task that also demands working memory could be the maintenance of retrieved propositions in the focus of attention. If grammatical encoding takes a substantial share of working memory, it is possible that the remaining resource is not sufficient to keep all retrieved propositions in the focus of attention, and hence some become irretrievable. Since encoding in L2 is demanding, one way of exploring the role of working memory in a higher-level cognitive activity is to create a condition that encourages the use working memory. In order to achieve this, the author capitalised on the linearisation problem, which might be compounded by using phrases whose order of propositions mismatch the order of their corresponding surface structure in a target language. However, the results obtained in Experiment 1 and Experiment 2 did not substantiate the prediction of poorer recall for incongruent discourses. A fuller model that took into account individual differences in working memory capacity actually removed the main effects and interaction. While it implies that working memory capacity can explain a considerable amount of variance in the data, it also suggests that memory maintenance and the memory recall in discourse interpreting were not sensitive to the factors of word order and translation direction.

One principle in language production that discounted the initial assumption of concurrent tasks in TL language production, is the incrementality of grammatical encoding. This principle has been conceived as strategic and task-dependent (F. Ferreira & Swets, 2002). By the principle of incrementality, a speaker does not encode a sentence in its entirety before he/she starts speaking. This implies that an interpreter would not retrieve the content of a whole discourse and then translate – the time pressure also prohibits delayed output. So, as has been mentioned earlier, interpreters might be like any other speakers, who talk while thinking how to continue. As there is no incentive to retrieve everything in one go, it was assumed that working memory may not be overloaded by the two tasks (memory maintenance and grammatical encoding). Instead, most resource may be available to speech planning and grammatical encoding in interpreting. When working memory is not stretched to its limit, participants' memory performance would be hardly affected.

The null effect in Experiment 3 of discourse paraphrasing as compared to the performance of verbatim recall weakened the account that paraphrasing involves deeper semantic analysis than verbatim recall does. Even if there were no controversy surrounding the account of levels of processing (see Baddeley, 1978 to see why it was controversial, also see Glenberg, Smith, & Green, 1977 for modified version of Levels of Processing), cautions should be in order about the stage at which *the processing* occurs in the paraphrasing task. If it is assumed to take place during comprehension and result in a richer encoding, could it be elaborated on which dimension the encoding becomes enriched? The fact that paraphrasing usually requires an expression using different word form or word order from its original seems to suggest that this dimension focuses on the surface form rather than deeper semantic encoding. And if the process leaves behind effective traces for memory retrieval, a paraphraser would need to retrieve the memory for the original form or word order in order to do something about it. This, however, will be contradicting the established finding that memory for surface code is short-lived (J. Bransford & Jeffery, 1971; J. D. Bransford, Barclay, & Franks, 1972; Murphy & Shapiro, 1994). Perhaps this is why interpreters' memory performance was essentially the same in paraphrasing congruent and incongruent discourses (Figure 3.4), i.e., they were biased to pay attention to the surface structure of discourses and encoding was shallower. But when they comprehended discourses in verbatim condition, qualitatively different comprehension processing from that in paraphrasing might make participants sensitive to prominent features that were more available in incongruent than congruent discourses, hence more coherent mental representation and better recall. The fact that participants in Experiment 3 only paraphrased half of all discourses in the paraphrase block suggests that it was difficult for them to follow this instruction. If paraphrasing was more difficult, participants would try to avoid it. However, the fact that twenty-five out of ninety-six discourses were paraphrased in verbatim block suggests that participants might not be aware of the difference in the surface form between original discourses and their recall.

The final experiment appeared to support the account of intrinsically different memorability between discourses that are congruent or incongruent in word order between Chinese and English. It also highlighted that comprehension is an important, if not the most important component of translation. This point was clearly demonstrated in Experiment 1 and Experiment 2 in their effects of translation direction: interpreters' perhaps better comprehension in L1 than L2 was associated with better memory performance in L1-L2 interpreting.

Although discourse interpreting in the four experiments provided informative data, it also showed some limitations. Its off-line measure (percentage recall) implies that quite a few assumptions have to be made in a complex task like language interpreting. Since a discourse is usually composed of several sentences, it can become very difficult to stringently control variables. For example, the discourses in Experiments 1-4 contained sentences that might vary in their syntactic complexity, information structure, and several types of modifiers that give rise to word order differences between English and Chinese. These were the reasons why there was a paradigm shift to the use of sentences and on-line measures in the rest of experiments in this thesis. The focus of the second series of experiments still was the role that working memory might play in language interpreting, but the locus that might implicate working memory was assumed to be comprehension by following findings by Macizo and Bajo (2006) and Ruiz et al. (2008). The basic research question remained the same: whether working memory plays a role in language interpreting, and if it does, how?

Chapter 4

Sentence Interpreting

4.1 Introduction

The question of how working memory plays its role in language interpreting still remained the overarching question. The approach taken in the series of experiments in this chapter was different from that in the last chapter. Apart from examining participants' memory performance in interpreting, on-line measures were employed in order to gain insight into the process of interpreting, assuming it is incremental and can be accommodated by major theories of language production. On-line measures were also expected to bring out the power of the approach that this thesis took throughout – a full consideration of the four translation components. The organisation of this chapter is similar to the last one. First, the general methodology used in this series of experiments is described. Then the research questions are formalised before each experiment is presented.

One issue with the recall-based methodology used in the last chapter has been the difficulty of relating results to hypotheses, especially as interpretation theories are not able to allow precise predictions. In the last chapter, available evidence did not substantiate the role that was assumed for working memory in discourse interpreting, as participants' memory performance was not sensitive to word order and translation directionality. As mentioned in the discussion, these results did not rule out a potential role that working memory had during language interpreting, because long-term memory could provide ample support, perhaps in every step of the process from SL comprehension to TL production. In a way, long-term memory confounded the design. That is why Baddeley et al. (2009) came up with a design for using 'constrained sentences' that could fall somewhere on a continuum of 'naturalness of speech': perfectly natural connected phrases at one extreme and unrelated word lists at the other. Their design was an attempt to reduce the support of long-term memory to its minimum. For a task like discourse interpreting that depends to a considerable extent on long-term memory, it might wash out any effect of factors that implicate working memory. As said earlier, memory performance might not be able to reflect the functional significance of working memory, which has been explored using on-line measures to gauge mental effort that fluctuates throughout a task. In order to address the issue in the last chapter, the self-paced reading paradigm in Ruiz et al. (2008) was adopted. The on-line measures will be complemented by measuring participants' memory performance in sentence interpreting. By using both on-line and off-line measures, it was hoped that full consideration of the translation component together with state-of-the-art psycholinguistic theories would shed some light on processing in language interpreting.

4.1.1 Revised hypothesis: Working memory as a workspace interfacing comprehension and production during SL comprehension

The findings that factors of word order and translation direction did not show their effect in discourse interpreting in the last chapter signalled that hypotheses had to be revised in further exploration. The locus where working memory modulated interpreting performance was assumed to be the stage of TL production. But as the review section showed, memory performance for discourse can be affected by recall methods and syntactic structures, therefore, perhaps working memory has some role to play during comprehension for later interpreting. At the same time, the null results in the last chapter should not be taken to trivialise the role of working memory in the production stage of interpreting only because there was no effect of word order. The hypothesis for the role of working memory in dealing with the linearisation problem in grammatical encoding remains valid, but it might need a change of methodology or hypothesis to test it.

One candidate method can be found in a few related self-paced reading time studies (Macizo & Bajo, 2006; Ruiz et al., 2008). These studies were motivated by one debate concerning whether there is an extra stage between comprehension and production in language interpreting and if there is, exactly when does it take place? This debate has developed into two major hypotheses of translation processing most specifically for translating under great time pressure such as SI (Danks & Griffin, 1997; De Groot, 1997). Assuming that there is no *recoding* phase, the sequential (or vertical) translation treats SL decoding and TL encoding entirely discrete processes. Translators only start encoding the to-be-translated message in a TL after source discourse comprehension is completed. Once the meaning is extracted from a source discourse, its surface form is discarded and only its meaning remains activated. Translators then reproduce the message in a TL on the basis of the activated meaning, as if it is a spontaneous language production. Therefore proponents of sequential translation hypothesis tend to think that the comprehension process in a translation task is no different from within-language comprehension.

In contrast to sequential translation, parallel (or horizontal) translation can be understood as incremental TL information activation at both lexical and syntactic levels (Gerver, 1976). Parallel translation also assumes that recoding could take place between languages during SL input analysis. Parallel translation assumes that TL encoding can overlap SL comprehension in time, therefore SL comprehension in preparation for later translation differs from within-language comprehension in terms of the resource that is required.

The third hypothesis is a hybrid of the two aforementioned translation hypotheses. In this hypothesis, translators can alternate between sequential and parallel translation depending on their competence and the way in which the source discourse is presented (De Groot, 1997). Parallel translation has been reported to be more often used by novice translators or when translators' performance breaks down temporarily (Seleskovitch, 1976; Paradis, 1994b). And Seleskovitch (1976) contended that parallel translation is more probable when the source discourse is permanently present for inspection, e.g., sight translation, but not when it is transient in a task such as SI.

Recently Macizo and Bajo (2006) directly tested the first two translation hypotheses and showed that interpreters as well as fluent bilinguals accessed TL lexical properties in a self-paced reading study. In two of their four experiments, participants were presented with Spanish sentences word-by-word, after which they were required to repeat sentences aloud in Spanish or translate them into English. Among their professional interpreters, there was a significant cognate effect, i.e., participants' reading time (RT) was shorter when the Spanish words were English cognates than when the words were not cognates. Crucially, the cognate effect was only present in reading-for-translation but not reading-for-repetition condition. The same results were replicated in a group of fluent Spanish-English bilinguals as well. These results were interpreted as evidence for parallel lexical activation in an interpreting context. And it was also supported by the cognate effect found in an ERP study with a similar design (Ibáñez et al., 2008). More recently, Ruiz et al. (2008) tested these translation hypotheses further by capitalising on a well-studied effect of lexical frequency in reading, i.e., low-frequency words tend to be fixated on or read longer (Rayner & Duffy, 1986). In their ingenious design, Spanish words that were matched for frequency were divided into high and low-frequency types according to the frequency of their English translation equivalents. Otherwise their procedure was similar to that of Macizo and Bajo (2006). Ruiz et al. replicated the effect of recall method, but most interestingly, there was a frequency effect as well. The reading time of Spanish words were longer when their English translation equivalents were low-frequency words than when they were high-frequency words. Since the frequency effect was only found in reading-for-translation condition, this interaction between recall method and lexical frequency provided further support for the parallel translation hypothesis. Macizo and Bajo suggested that the longer RT may have resulted from an extra cost the parallel translation incurred during TL lexical retrieval over and above the normal resource that is demanded for monolingual comprehension.

More recently, Ruiz et al. (2008) extended their inquiry of TL activation during translation to the syntactic level. In their second self-paced reading experiment, the flexible word order of Spanish adjectival phrases was used to test the translation hypotheses. Compared to the rigid Adj-Noun (A-N) order of English adjectival phrases, Spanish adjectives can be placed on either side of the nouns they modify, although there is a subtle pragmatic difference between the two usages. Each of Ruiz et al.'s Spanish sentences had two versions which differed only in the word order of the adjectival phrase. The congruent version had the same word order as its English translation (e.g. verde césped - green lawn), whereas the incongruent version reversed the adjective and the noun (e.g. césped verde - lawn green), hence a mismatch of word order. Their rationale was that if translation involves parallel TL syntactic activation, reading times would be longer in incongruent than congruent conditions, and this reading time difference would be present in the translation condition only. On the contrary, if reading for translation did not implicate parallel TL syntactic activation, there would be no difference in reading times. Their results confirmed the effect of recall method where the reading time in reading-for-translation conditions was reliably longer than that in reading-for-repetition, suggesting that resources had been allocated in accessing TL lexical information. Most importantly, their participants were slower in reading when word orders were incongruent than when they were congruent, but only in the reading-for-translation, not the reading-for-repetition condition. This interaction led Ruiz et al. to conclude that the congruency effect may be driven largely by the activation of TL syntax and perhaps also by searching for syntactic matches in the TL, both of which demand working memory. The effect of word order congruency between SL and TL has also been reported in a study using eyetracking (Sjørup et al., 2009) in which professional interpreters read Dutch sentences for later translation into English while their eye movements were recorded. Compared to the fixation duration on Dutch phrases with subject-verb word order, longer fixation duration was observed before their participants translated phrases with verb-subject word order, which is incongruent with the subject-verb order of its English translation.

In this brief review, evidence shows that TL can be activated during comprehension for translation at both lexical and syntactic levels. But is there a role for working memory in the process of reading for translation as suggested by Macizo and Bajo (2006)?

Traditionally, the role of working memory in language processing can be explored with two approaches. The correlational approach relates participants' working memory measures to their performance of the tasks in focus. Christoffels et al. (2003) reported a significant correlation between English(L2) reading span and simultaneous interpreting performance (judged by professional interpreters on how well sentences were translated) among a group of untrained Dutch-English bilinguals. Alternatively, the dual-task paradigm has been widely used in investigating the role of working memory in language comprehension. In this paradigm, the primary task is usually an on-line reading task followed by sentence verification, and the choice of a secondary task depends on the investigators' hypothesis of the mechanism underlying language comprehension (Fedorenko, Gibson, & Rohde, 2006, 2007). By applying additive factors logic (McClelland, 1979), if sentence processing and memory tasks share the same resource pool, the manipulated factors should interact, otherwise the reaction time should show a strict additive effect if two tasks draw on different resources.

By the same token, this paradigm can be used to test the revised hypothesis of the role of working memory in language interpreting. According to Baddeley and Hitch (1974b), working memory can be thought as a system consisting of a limited capacity 'work space' which can be divided between storage and control processing demands. With the evidence provided by Macizo and Bajo (2006) and Ruiz et al. (2008), it was hypothesised that working memory served as a workspace that interfaced SL and TL lexicons by not only recognising SL lexical items but also activating TL information stored in long-term memory. Following Ruiz et al. (2008), TL activation might include searches for lexical items and syntactic matches, which could be resource demanding. The following section briefly describes the general methodology used in Experiments 5-8.

4.1.2 General methodology

The on-line measure used in this chapter was a self-paced reading time task. The dependent variable was the reading time for individual word or area of interest. The rationale of using this kind of methodology is based on an old assumption that "processes that require more attentional or memory resources take longer than processes that require fewer resources" (Zwaan & Singer, 2003, p. 87). Ruiz et al. (2008) required participants to read sentences by pressing a button repeatedly to show a sentence word-by-word at the centre of a computer screen. One deviation in the present procedure from the paradigm used by Ruiz et al. (2008) was the use of a moving-window presentation. It was thought that by informing the participants of the length of sentences, it would lower the possibility of participants' inconsistent reading behaviour for not knowing the length of each sentence. For instance, when the length of a sentence was unknown and participants feared that their memory for the earlier part of a sentence could decay before they reach the end, they might read faster than they usually do in order to quickly reach the end of a sentence. In addition, participants were encouraged to read sentences at their natural reading speed.

Another deviation of procedure from that of Ruiz et al. (2008) was the introduction of a secondary task in Experiments 6 and 7. The secondary task was digit recall, with digits preloaded before participants started to read sentences.

As for the material, Experiment 5-7 used the same set of sentences and Experiment 8 used another set. Since the material was not the same for all experiments and the procedure was slightly modified for Experiment 6-8 on the basis of that of Experiment 5, these details are described in the methodology section of each experiment. The scoring procedure was the same as described in Chapter 3.

4.2 Experiment 5 – A replication of Ruiz, Paredes, Macizo, and Bajo (2008)

The objective of Experiment 5 was to replicate the congruency effect observed by Ruiz et al. (2008) in which reading time was longer for phrases whose SL structure differed from their TL in a no-load condition. This was deemed important before exploring the role of working memory by using the dual-task paradigm.

Contrary to sequential translation which posits that the comprehension process does not differ between reading for translation and reading for comprehension, parallel translation assumes that during reading for comprehension, target language lexical and syntactic properties are activated. The target language activation was assumed to be modulated by working memory and it was expected that it would lead to a longer reading time in reading-for-translation than that in reading-for-comprehension conditions, i.e., an effect of recall method. The working hypothesis also predicted that when source and target language differ in word order, this difference could trigger an active search for syntactic matches, and lead to a congruency effect, whereby phrases that have different word order in SL and TL will be read slower than those which did not differ. Finally, recall method and word order would interact.

4.2.1 Method

Participants

Sixteen (3 males and 13 females) native Chinese speakers from Taiwan who were advanced learners of English took part in this experiment. These participants were all postgraduate students recruited from the University of Edinburgh and their IELTS (International English Language Testing System) scores at the time of admission ranged between 6.5 and 7.5 (M= 6.8, SD= .3). Each of them was paid six pounds for this 1.5-hour experiment.

Material

Forty-eight sentences were selected from four studies (Fedorenko et al., 2006, 2007; Hsiao & Gibson, 2003; Lin & Bever, 2006) whose materials contained object-extracted relative clauses (see section 3.1.2 for the word order differences between English and Chinese). The first step of material preparation was to acquire both Chinese and English versions of all sentences. Hsiao and Gibson (2003) and Lin and Bever (2006) had supplied both versions of sentences in their articles. Those sentences that were selected from Fedorenko et al. (2006, 2007) were translated from English into Chinese by the author who is a trained interpreter. When this first step was completed, these sentences were then subjected to word frequency control. Since word frequency affects sentence reading time (I. M. Liu, Wu, & Chou, 1996; Rayner & Duffy, 1986; Seidenberg, 1985), the frequency of each Chinese word in the first clause of each sentence was checked and replacements were made when necessary in order to make sure that none of the words in the first clause of each sentence were low-frequency words. The area of interest in each sentence was the third region of the first clause: NP(animate subject)+ relative clause(transitive verb)+ NP(animate object). Because it has been suggested that the frequency of a target-language equivalent seems to influence RT at the sentence-final, but not the sentence-initial position in the reading-for-translation condition (Ruiz et al., 2008), it was decided that no word replacement was necessary as long as their English equivalents were not low-frequency words. By low-frequency words, it is meant to the words whose frequency counts are lower than a critical value, which were 2.1/million for Chinese (Y. Liu & Perfetti, 2003) and 5/million for English (Kućera & Francis, 1967). The average number of characters in a Chinese sentence was 16 (SD = 1.7), and the average number of propositions was 4 (SD= 1). When the second step of material preparation was completed, a complete list of 48 sentences that contain object-extracted relative clauses were used as the incongruent version of the experimental sentences.

In order to construct the congruent version for each of the forty-eight incongruent sentences with minimal changes to their meaning, the Chinese relativiser 'DE(的)' was removed from the first clause of each incongruent sentence. The removal of 'DE(的)' turned a matrix sentence into a primary clause followed by a verb phrase. The primary clause (NP-VP-NP) could be translated in such a way that the word orders of Chinese and English did not differ, so that the word order of SL text was congruent with that of its TL text. However, the removal of 'DE(的)' from a matrix sentence resulted in an ungrammatical sentence in both Chinese and its English translation, because a primary clause would be followed immediately by a verb, e.g., 護士協助醫生借了手術的器具 (The nurse assisted the doctor borrowed equipments for the operation). Therefore a pronoun 'he(他)' or 'she(她)' was inserted after the primary clause to act as the subject of the following clause. All the congruent sentences were generated by applying this procedure to each incongruent sentence (see below for example sentences).

28. Incongruent:
Chinese:護士協助的醫生借了手術的器具
English:The doctor(醫生) who the nurse assisted(護士協助的) borrowed(借了) equipments for the operation(手術的器具).
29. Congruent:
Chinese:護士協助醫生。他借了手術的器具
English:The nurse(護士) assisted(協助) the doctor(醫生). He(他) borrowed(借了) equipments for the operation(手術的器具).

Another forty-eight sentences with different structures other than those in experimental sentences were generated and used as fillers. Four sentence lists were developed to achieve counter balancingby using Latin Square design. In each list, half of the sentences were congruent sentences, and the other half were incongruent. A participant would only see one version of each sentence, but both versions of sentences would be seen in reading-for-comprehension and reading-for-translation conditions.

Procedure

Participants read Chinese instructions on the screen and started the practice session when they had no questions about the experimental procedure. Each participant had to complete two blocks of 96 trials, half of which were fillers. The order of blocking was counterbalanced and the sentences within each block were randomised. Half of the participants started with the repeat block, and the other half with the translation block. Experimental sentences were separated by a filler sentence. Linger (Rohde, 2009) was used to present testing material in the moving-window mode and record participants' reading times.

A typical trial started by presenting a line consisting of a number of short lines, each of which masked a Chinese character. So the sentence length was clearly indicated by the length of the entire line. Each sentence was vertically centred and horizontally aligned to the left of the computer screen. Participants were instructed to start reading as soon as they saw the line by pressing the space bar of their keyboard and to continue doing so until they finished reading the whole sentence. Each press of the space bar would reveal one Chinese word. For instance, if it was a two-character Chinese word, participants would see two characters at the same time, rather than one character upon each key stroke. At the end of sentence presentation, participants were prompted to carry out their task by a cue on the centre of the screen 'Translate Now' in translation block or 'Repeat Now' in repeat block. The visual cue was accompanied by a chime played through a set of speakers. In the reading-for-repetition block, participants had to repeat aloud as accurately as possible immediately after they finished reading a sentence. In the reading-for-translation block, they had to start translating a sentence aloud from Chinese into English. Their utterances were recorded by Audacity (Version 1.26) for analysis. After the reading task, each participant had to complete a Chinese reading span task.

4.2.2 Results

The main data analysis was conducted on two measures, one was participants' reading time at the area of interest, which was the third region (object NP) of the first clause in each sentence, and the other was their memory performance in terms of the percentage of propositions correctly reproduced in their sentence repetition and translation tasks. The dataset was trimmed in two stages. Extreme data points shown on the individual's boxplot were removed. And then for each participant, reading times that were 2.5 SD or above the individual mean were replaced by the cut-off value, i.e., mean + 2.5 SD.

These data was then subjected to an ANOVA with reading times of the critical words as the dependent variable, while the recall method and word order congruency were the independent variables. The counterbalanced group was the between subject variable.

The ANOVA showed a main effect of recall method in by-subject analysis, $F(1,15) = 15.5$, $p<.05$; reading for translation (1728 ms) was significantly longer than reading for repetition (886 ms). This effect was also significant by item, $F(1,46)=67$, $p<.01$). Reading times were reliably longer in incongruent (1446 ms) than congruent condition (1168 ms) in by-subject analysis, $F(1,15)=5.6$, $p<.05$, but was only marginally significant by item, $F(1,47)=3.53$, $p=.05$. Crucially, word order congruency and recall method interacted. Figure 4.1 shows that reading incongruent sentences was slower than reading congruent sentences in translation, but not repetition conditions (see Table A.4 in the Appendix for condition *M* and *SD* values.). This interaction was significant by subject, $F(1,15)=5.7$, $p<.05$), but it did not reach significance in by-item analysis. None of the factors interacted with the between-subject variable, counterbalanced group.

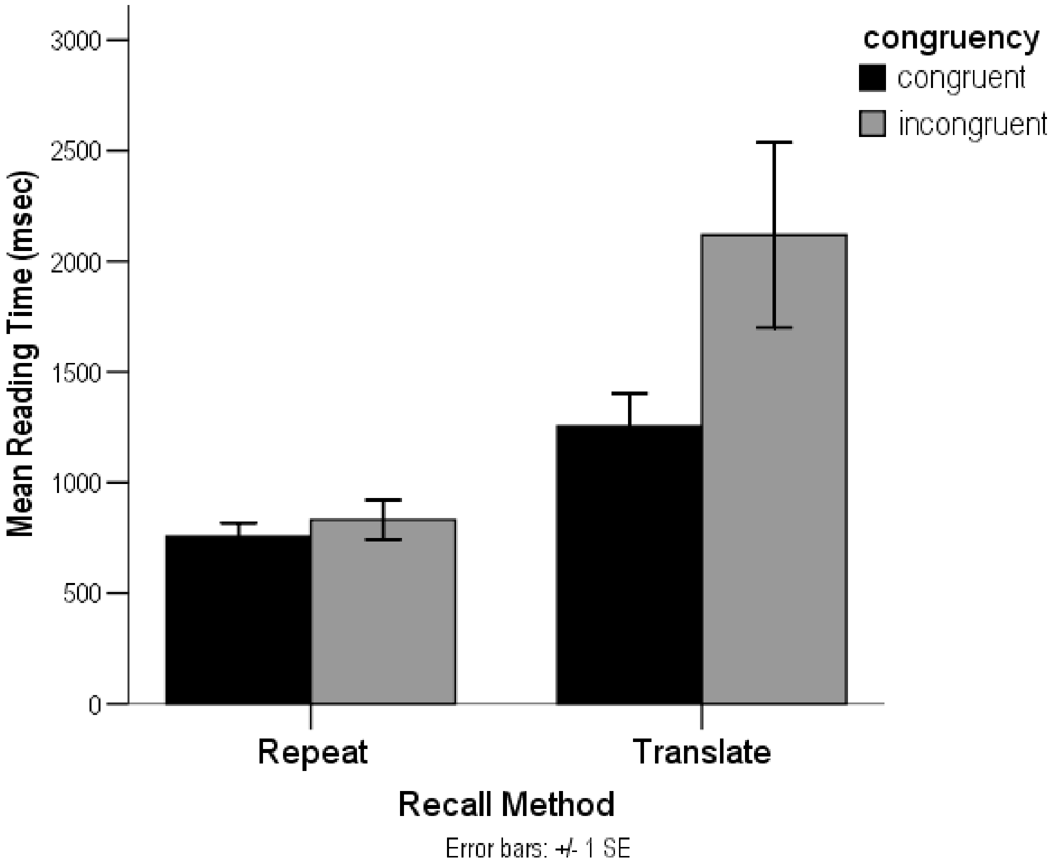


Figure 4.1: Reading times as a function of recall method and word order congruency.

For the analysis of participants' production data, participants' responses were transcribed and scored against a textbase template using the same procedure as in the first four experiments. Figure 4.2 shows the mean percentage of propositions reproduced in each condition (see Table A.5 in Appendix A for M and SD values). A mixed ANOVA was conducted using the mean percentage of correctly reproduced propositions as the dependent variable, the word order and recall method as independent variables and the counterbalanced group as between-subject variable. In by-subject analysis, there was only one significant effect of recall method: memory recall was superior in reading-for-repetition (96%) than reading-for-translation conditions (89%), $F(1,12)=23.2$, $p<.01$. In by-item analysis, again, there was an effect of recall method on memory: repetition condition resulted in better recall than the translation condition, $F(1, 46)=19$, $p<.01$.

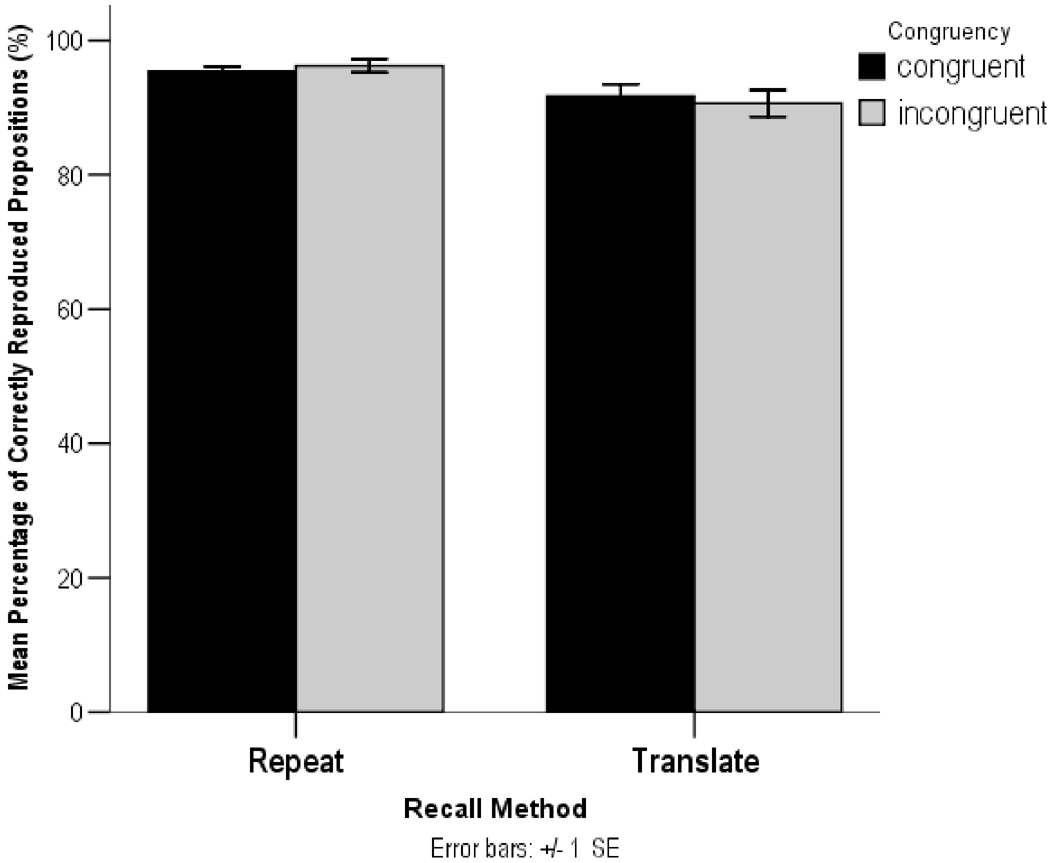


Figure 4.2: Reading times as a function of recall method and word order congruency.

4.2.3 Discussion

The reading time data replicated the results of Ruiz et al. (2008). Participants were slower in reading incongruent sentences than congruent sentences only when they were required to translate but not when they were asked to repeat them after reading. This result was consistent with the prediction of the parallel translation hypothesis which suggested that translators can actively access the syntactic properties of a TL at the same time when they are analysing the source language input. It was suggested by Charles Clifton in a local conference (Jin, 2008) that the longer reading time in translation conditions could be driven by the fact that relative clauses are intrinsically more difficult to read than simple clauses. If this is the case, longer reading times for incongruent sentences should be found in the repetition condition as well. But the results showed that the reading times did not differ between congruent and incongruent sentences in reading-for-repetition conditions. The other evidence that could argue against the criticism can be found in the online sentence survey. In a post-hoc test, experimental sentences were rated on their syntactic naturalness and plausibility in an web-based survey. It was not a paid survey, but those participants who completed the entire survey were eligible for a prize draw of five scarves. 110 participants recruited in three universities in Taiwan took part, but only 48 (14 males and 34 females) of them completed the entire survey. By naturalness, each sentence was rated on a 1-to-7 scale as to how natural it sounded to a native Chinese speaker. In other words, survey participants were asked how likely it was that a native Chinese speaker would produce sentences in the survey. By plausibility, participants were instructed to rate each sentence on a 1-to-7 scale as to how likely it was that the description in each sentence could happen in reality, irrespective of its naturalness. A t-tests confirmed that incongruent sentences were rated higher than congruent sentences in both their naturalness, $t(47) = 2.5$, $p<.01$, and their plausibility, $t(47)=4.5$, $p<.01$. Given that incongruent sentences were rated higher on their syntactic naturalness, it would facilitate reading, and result in shorter rather than longer reading times. It was therefore concluded that the effect of recall method seems a result of target language activation during source language comprehension for later translation. The interaction between recall method and congruency indicated that word order difference between source and target languages plays a key role in the ease of syntactic processing for later production in a target language during translation. Participants' higher proportion of content reproduction in sentence repetition than sentence translation showed that translation had an impact on the completeness of sentence reproduction. This result is compatible with the findings of Macizo and Bajo (2006), who found that production quality was poorer in translation condition than in repetition condition.

Ruiz et al. (2008) suggested that longer reading time for incongruent sentences than congruent sentences was driven by participants' searching for TL syntactic structures. While this possibility remains attractive and reasonable, there did not seem to be solid evidence to associate the reading time difference and searching for TL syntactic structure. What might be more tractable in further exploring the findings in the present experiment could be testing how working memory was involved in comprehension for translation. To be more specific, it can be asked which function of working memory was involved in the process that led to a reading time of as long as two seconds on one Chinese word before translating a given sentence into English. In the next experiment, the dual-task paradigm was adopted to address this question.

4.3 Experiment 6 – Exploring Working Memory using the dual-task paradigm

In a dual-task study, Miyake, Just, and Carpenter (1994) established that low-span participants had more difficulties in maintaining multiple representations of an ambiguous word (homograph) than their high-span counterparts. But this ambiguity effect, manifested by longer total RT of a disambiguating phrase, was only significant when participants read sentences in which the distance between an ambiguous word and its disambiguating word was long (i.e., separated by up to 7 intervening words), but not when it was short (see below), indicating that low-span individuals were more sensitive to cognitive load increase.

High load: Since Ken liked the *boxer* very much, he went to the nearest *pet* store to buy the animal.
Low load: Since Ken liked the *boxer*, he went to the *pet* store to buy the animal.

By adopting the similar reasoning and design as that of Miyake, Just, and Carpenter (1994), Macizo and Bajo (2006) recorded 16 translators' reading times when they read different types of Spanish sentences (ambiguous vs. unambiguous) for later repetition or translation into Spanish. In ambiguous sentences, they manipulated the memory load by inserting five words between an ambiguous word and its disambiguating word for the low-load and seven words for the high-load version of sentences. As predicted, the reading times of the disambiguating words were longer in reading ambiguous sentences than that for unambiguous controls and the ambiguity effect was only significant in the high-load, but not in the low-load condition. Critically, the ambiguity effect was present in reading-for-translation, but not reading-for-repetition condition. Their off-line measures suggested that ambiguity and load had a clear impact on their sentence verification task and production quality. The sentences with ambiguous words and higher load led to poorer accuracy than control sentences in the sentence verification task. Although the authors did not report the analysis of the production quality, a t-test using supplied data in the article confirmed that the production quality (rated on a 1-7 scale) was poorer in translation (M=4.79, SD=0.95) than that in repetition (M=6.37, SD=0.86). Macizo and Bajo (2006) showed that holding multiple representations of an ambiguous word across a stretch of time was costly and it became even more costly when translation was required after sentence reading. Along the same vein, the present experiment was designed to address the same question, but with a different approach.

The dual-task technique was used to explore what function of working memory might be implicated in the reading for translation task. Hypothesis framing regarding exactly what mechanism it was that taxed more resource and led to longer reading time was avoided. If, for instance, the phonological system modulated the process that demanded extra resource and led to the reading time difference between congruent and incongruent sentences, a secondary task that uses up the same resource might cost and make the processing much longer than when there is only one task. In one experiment, the phonological system was assumed to be the modulating component of working memory, therefore digit preload was chosen as the secondary task. The working hypothesis was that if the activation of TL syntactic properties in reading for translation tapped into the same resource as did digit memory, it would show that the higher the digit preload, the less resource is available for the activation of TL properties, and therefore it would take the participants much longer to access the TL information. By following the additive logic, the higher the digit preload, the longer the reading times, and the larger the reading time difference between congruent and incongruent sentences. Alternatively, digit preload might be so demanding that it could indiscriminately use up participants' resource, leaving them with too little capacity to access TL information. In this case, word order congruence and digit preload would not interact.

4.3.1 Method

Participants

Eighteen participants (3 males, 15 females) who were native Chinese speakers were recruited in the University of Edinburgh for this experiment. They were aged between 23 and 25 (M= 23.7). Each participant was paid 10 pounds for this 1.5-hour experiment. They all had IELTS scores of between 6.5 and 7 (M= 6.8, SD= 0.3). None of these participants had taken part in Experiment 5.

Material

The sentences in this experiment were the same set as used in Experiment 5.

Procedure

The procedure only differs from that of Experiment 5 in the addition of a secondary task in the present experiment. The secondary task was digit recall. Since this could be the first attempt of using digit preload in a reading-for-translation task, there was no clear indication of how large the preload must be in order to bring out the effect. Considering that Chinese speakers are well known for their superiority in digit span due to the monosyllabic numericals (Hoosain, 1984), the high-load condition used a 7-digit-list and the low-load condition used a 5-digit-list. There was a no-load condition as the control condition. So the three levels of preload together with two levels of word order congruency constituted a three-by-two repeated measure design.

The experiment was conducted individually in a sound-proof computer lab. Participants read instructions on the computer screen and started practice trials before conducting the real test. Each participant had to complete 96 trials, half of which used fillers. As there were six conditions, there were eight data points for each condition. Because the research condition concerned the role of working memory in sentence translation, i.e., how participants reacted to the increase of preload when at the same time they were attempting to access TL information during comprehension, the dependent variable for all trials in this experiment was translation. The procedure remained exactly the same in the no-load condition: participants read either congruent or incongruent sentences for translation. In the load conditions, however, they were first presented with lists of 5 or 7 digits, all digits simultaneously presented on the screen. The presentation time was the multiplication of 300(ms) and number of digits in a list. After the presentation of digit list, the self-paced reading began. Once a participant finished reading the whole sentence, a box appeared on the centre of the screen for recall of the digit list. This box did not appear alongside an instruction (but participants had done practice trials to familiarise themselves with the procedure). Therefore, a participant had to immediately interpret the sentence. Experimental sentences were pre-randomised and separated by fillers. As usual, participants' responses were recorded digitally for analysis, and the software automatically logged participants' digit recall. The scoring for translation production followed the same procedure as described in Experiment 1.

4.3.2 Results

As this is a dual-task paradigm, there were two dependent variables: reading time and digit recall. The independent variables were word order congruency and digit preload. Reading time data was screened by replacing extreme data points with the cut-off value which was the mean plus 2.5 standard deviation.

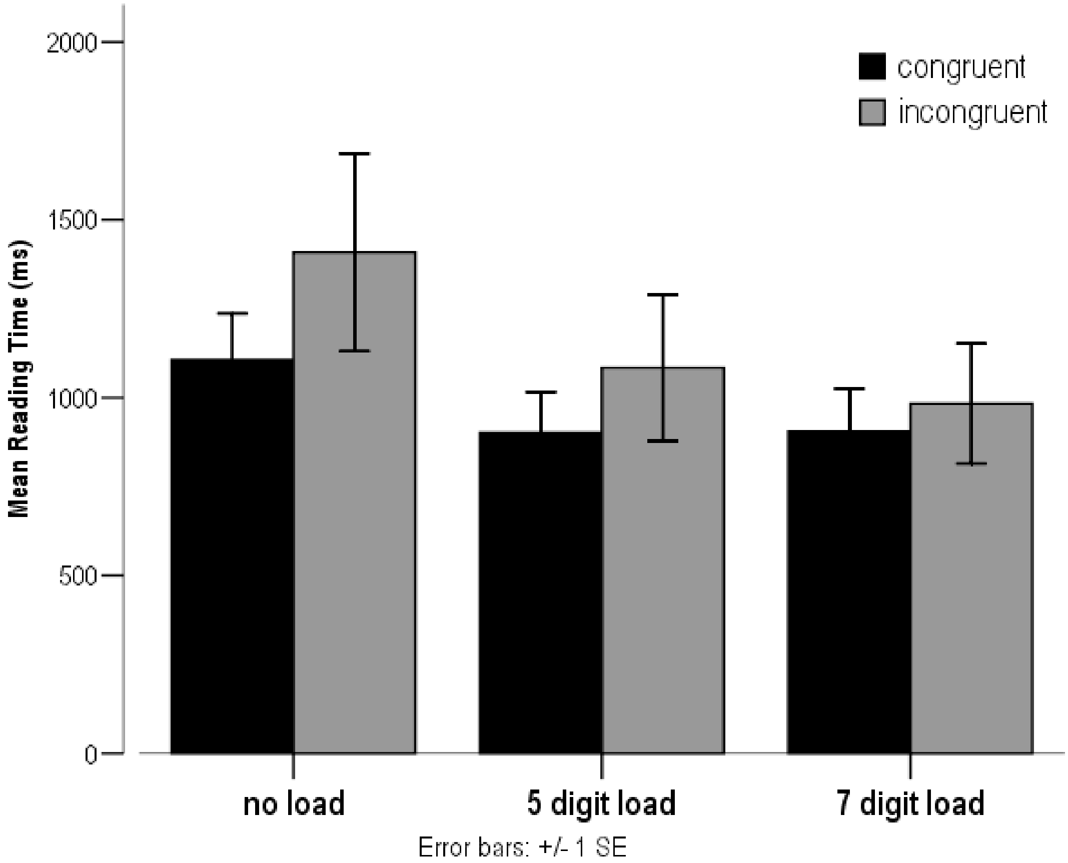


Figure 4.3: Reading times as a function of digit preload and word order congruency.

Figure 4.3 shows that mean reading time decreased as digit preload increased (see Table A.6 in the Appendix for condition *M* and *SD* values). A repeated measures ANOVA was conducted using digit preload and word order congruency as independent factors and the mean reading time as the dependent variable.

In by-subject analysis, there was an effect of digit preload, $F(1,17)=9.2$, $p<.01$. A post-hoc test established that the effect of preload was driven by the reading-time difference between high-load and no-load. Nothing else reached significance. Post-hoc power analyses confirmed that the observed power (assuming that population effect size is equal to the sample effect size) for within-factor effect of word order congruence, digit preload, and their interaction was 0.31, 0.87, and 0.3 respectively, suggesting that sample size could increase in future replications or in fact, future studies might re-consider their working hypotheses and experimental design.

In by-item analysis, except the effect of preload, $F(1,47)=5.3$, $p<.05$, no other effect or interaction reached significance. A main effect test indicated that the preload effect was driven by the significant reading time difference between no-load and 5-digit load conditions.

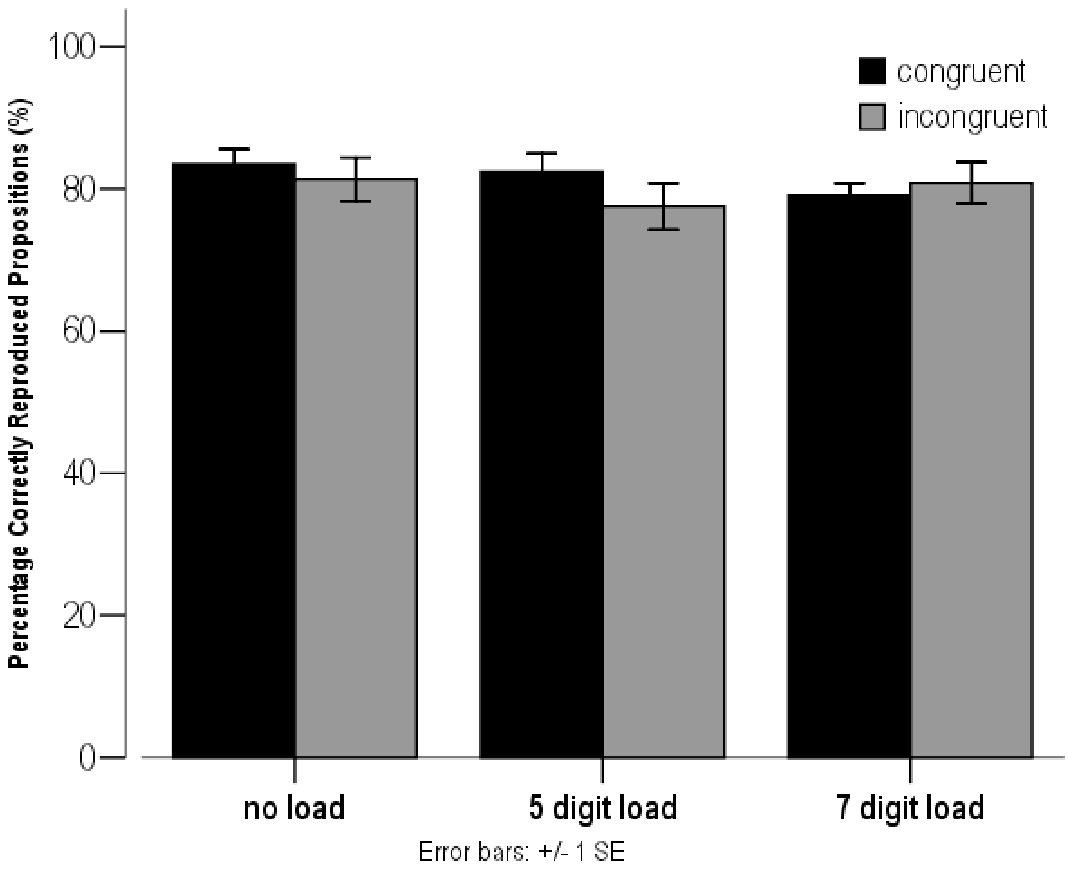


Figure 4.4: Mean percentage of correctly reproduced propositions in sentence interpreting in all conditions.

Participants' memory performance in sentence interpreting was also examined. Neither by-subject nor by-item analysis yielded significant effects or interactions (4.4). It appeared that participants' sentence memory was not sensitive to word order congruency or digit preload.

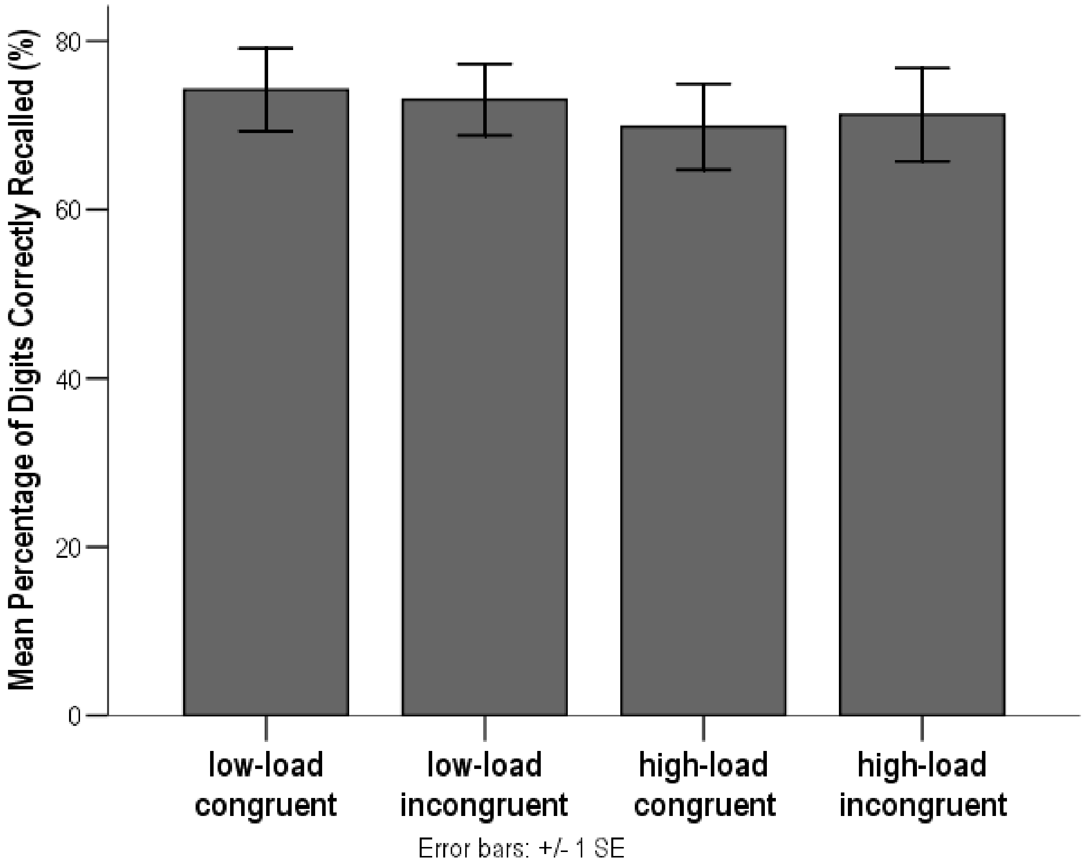


Figure 4.5: Mean percentage of correctly recalled digits in preload conditions.

Finally, digit recall in preload conditions showed no effect of preload or word order congruency in a repeated measures ANOVA (Figure 4.5), suggesting that digit memory was as good in five-digit as seven-digit conditions.

Thus far, analysis of reading times and memory for sentences as well as for digit lists has been presented. But this is not the end of story. During response transcription and scoring, participants' TL production demonstrated structure alternation between passive and active voices in translating incongruent sentences. For instance, participants' translation for the Chinese sentence 間諜認得的小偷正在紅綠燈過馬路 alternated between English translation A and B:

A *The thief who was recognised by the spy was crossing the road at the traffic light.*
B *The thief who the spy recognised was crossing the road at the traffic light.*

The question thus arose as to why participants did not stick to one type of construction and why some participants showed more alternation than others. In order to explore whether the construction choice was related to any of the factors, participants' responses were coded as active or passive for each trial, and then the percentage of active voice construction was obtained for each condition. The lower this measure was for a given condition, the higher the tendency it showed for someone to use the passive construction. As congruent sentences contained S-V-O word order in the first clause, it was predicted that participants would use active voice more often than the passive construction in sentence interpreting. However, there was no prediction for the frequency of passive construction in translating incongruent sentences.

An ANOVA confirmed the effect of congruency, $F(1,17)=22.8$, $p<.01$, suggesting that participants were significantly more consistent in using active voice in translating congruent sentences than in translating incongruent sentences (Figure 4.6). There was no effect of digit preload and the two factors did not interact.

4.3.3 Discussion

In the revised hypothesis, the role of working memory is conceptualised as a workspace where a coherent representation is constructed for a sentence and TL properties are accessed. If accessing TL information tapped into the same resource as digit memorisation did, one would expect to observe much longer processing time when

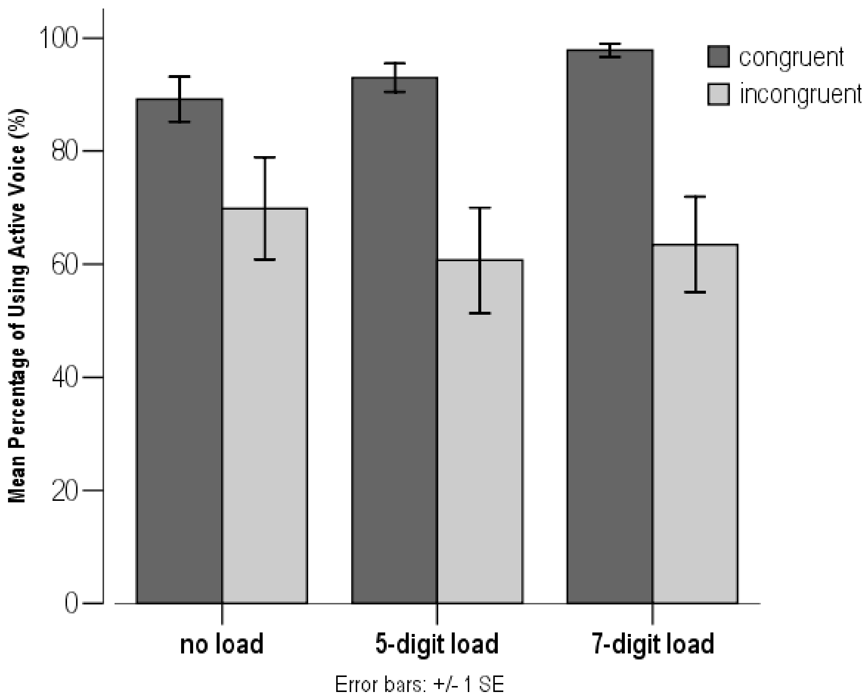


Figure 4.6: Proportion of active voice used in sentence interpreting as a function of word order congruency.

participants' working memory was occupied by digit preload. It would also be predicted that word order would interact with digit preload, i.e., reading times would be differentially affected by word order congruency depending on the preload. The data, however, told quite another story. Firstly, the experiment did not replicate the findings in Experiment 5, in which reading times were longer for incongruent sentences than congruent sentences when there was no digit preload. The most surprising result was that the load effect showed exactly the opposite pattern from the prediction. The reading times decreased as the preload increased. This intriguing pattern may be associated with the task demand. Since the task required participants to memorise digits and recall immediately after reading a sentence in load conditions, it was possible that participants prioritised digit recall over the parallel translation. In order to deliver the best performance in digit recall, they either had to refresh the list by active rehearsal of digits in their phonological working memory regularly, or they had to reduce the time lapse between encoding and retrieval, so that digit memory would not decay. One way of achieving the latter goal would be speeding up their reading, which would result in shorter reading time for critical words of sentences. Then, is it possible that participants might be less prone to prioritise digit recall if the order of task responses were reversed, i.e., translate a sentence and then recall the preload digits? This possibility was tested in Experiment 7.

One significant finding was that participants performed equally well in digit recall and sentence interpreting regardless of digit preload. It appears that participants could tune up their processing so that they could manage to do more things in the same period of time under load than when there was no load. The observation that faster reading did not compromise memory for sentences and digits seems to indicate that the functions of processing and storage of working memory are modular, which is more consistent with the multi-component working memory (Baddeley, 2000) and SSIR theory (Waters & Caplan, 1996b).

Coincidently, Macizo and Bajo's (2006) data also has an intriguing pattern, showing that the reading times of the disambiguating words in control sentences was numerically shorter in high-load than in low-load condition when translation was required. Recall that their high-load condition refers to more words being inserted between the target and disambiguating words, thus their participants might be using a reading strategy which is similar to that used by participants in the present experiment in order to maintain the alternative semantic representations of a target word in their working memory. One aspect of the procedure that could weaken this account, however, is that their participants should have been unaware of the distance between a target word and its disambiguating word, since sentences were presented word-by-word at the centre of the screen.

So what about the critical research question concerning the role of working memory in sentence interpreting? If the reading time difference between reading incongruent and congruent sentences was related to more resource than was demanded by searches for syntactic structures in a TL in Experiment 5, the failure of replication of this effect in Experiment 6 seems to argue for two possible accounts: 1) the mechanism underlying the reading time difference in no-load condition in Experiment 5 was not an absolutely necessary procedure for successful interpreting, as the interpreting performance in terms of sentence memory was not sensitive to load and word order congruency in Experiment 6. In other words, the mechanism could be strategic in nature; 2) extending from the first point, it appears that whether or not participants carried out the extra process depended on the resource available to them. It is possible that the demand of digit memory might have taken up too much resource of participants' limited capacity, and therefore the available resource left to the participants was only enough for the comprehension of individual word but not enough for any extra process, e.g. searches for syntactic structures, resulting in the absence of the word order effect. The results in this experiment therefore appear to support the alternative hypothesis. As accessing to TL information might be very resource-consuming, a secondary task that required attentional control took precedence over TL access during comprehension, preventing participants from accessing TL information on-line, hence the disappearance of word-order effect.

In order to address the issues of task prioritisation of one over the other and to create a condition that encouraged participants to carry out parallel translation, Experiment 7 reversed the order of task responses, so that participants translated sentences before recalling the preloaded digits. As there was essentially no reading-time difference between 5-digit and 7-digit preload conditions in the present experiment, the 7-digit preload condition was removed in Experiment 7. So the objectives of Experiment 7 were 1) to replicate the key finding of Experiment 5, the word order congruency effect; and 2) to investigate the potential advantage of congruent word orders between SL and TL. Since it has been reported that structural priming in sentence production was associated with shorter sentence onset latencies (Corley &

Scheepers, 2002; Smith & Wheeldon, 2001), it might be possible to observe a word order effect in the sentence onset latency. If participants were faster in sentence reading, it was possible that they would be slower in initiating interpreting since they did not have time to access TL information during comprehension; and 3) to continue observing the voice alternation in participants' relative clause construction. Even though the voice preference was not sensitive to load in the present experiment, it would be interesting to see whether the same result generalised to Experiment 7 where TL sentence production preceded digit recall. The implication of this setting was that while producing a TL sentence, participants might need to hold the digits for later recall, creating a dual-task condition similar to that in the comprehension phase. Thus, different performance might be observed in their sentence interpretation, digit recall, and voice preference in the construction of the first clause of each sentence.

4.4 Experiment 7 – A replication of Experiment 6 with small modifications

Experiment 7 was designed to explore exactly the same research question of Experiment 6: how working memory was implicated in the word order effect found in Experiment 5. Two major modifications were made to the procedure in Experiment 6 in order to address methodological issues that might have prevented participants from behaving the way the participants read for translation in Experiment 5. One new measure, speech onset latency, was introduced in the present experiment to see whether faster sentence reading would be associated with longer onset latencies in TL production.

4.4.1 Method

Participants

Sixteen Chinese-English late bilinguals (4 males and 12 females) were recruited for this paid experiment (6 pounds each). They were all postgraduate students aged between 23 and 29 (M=23.7) in the University of Edinburgh. They were originally from the People's Republic of China, and they had IELTS scores of 6.5 or above (M=7.1, SD=0.3). The material in the present experiment was the same set of sentences used in Experiment 5. None of them had participated in Experiment 5 or 6.

Procedure

This experiment used a two-by-two repeated measures design. Participants self-paced read congruent and incongruent sentences in no-load and 5-digit-preload conditions for later translation at their own pace. The procedure of sentence presentation was the same as in Experiment 5 and 6. The difference was that participants were required to translate Chinese sentences into English immediately after they finished reading them. In the no-load condition, the experiment automatically presented a new sentence when they pressed any key after translating a sentence. In the 5-digit-load condition, they first saw the digit list, and then read a to-be-translated sentence. After translating the sentence, they had to recall the digits by typing them in a blank box at the centre of a computer screen. Their verbal responses were recorded digitally for analysis. Because a chime was played through a pair of speakers at the end of each sentence, it was possible to obtain participants' voice onset latencies using the software Praat. Remijsen's script which is available for download on his website was adopted for this purpose (Remijsen, 2009). The criterion for the voice onset ignored the false starts, which were often followed by long pauses. The speech onset latency was the duration between the end of sentence presentation and the speech onset in sentence interpreting. In other words, this duration was the silent interval between two voices, the beep at the end of sentence and the utterance of a participant. The silent interval for a trial was calculated by the script by subtracting the reading of time stamp for participants' utterance by the time stamp for the beep. Each trial was examined to ensure that the script did not pick up background noise during the *silent* interval and did not produce more than one silent interval. A typical silent interval can be examined by opening a window that shows the sound wave and tiers (see Figure 4.7). The response scoring followed the procedure described in Experiment 1.

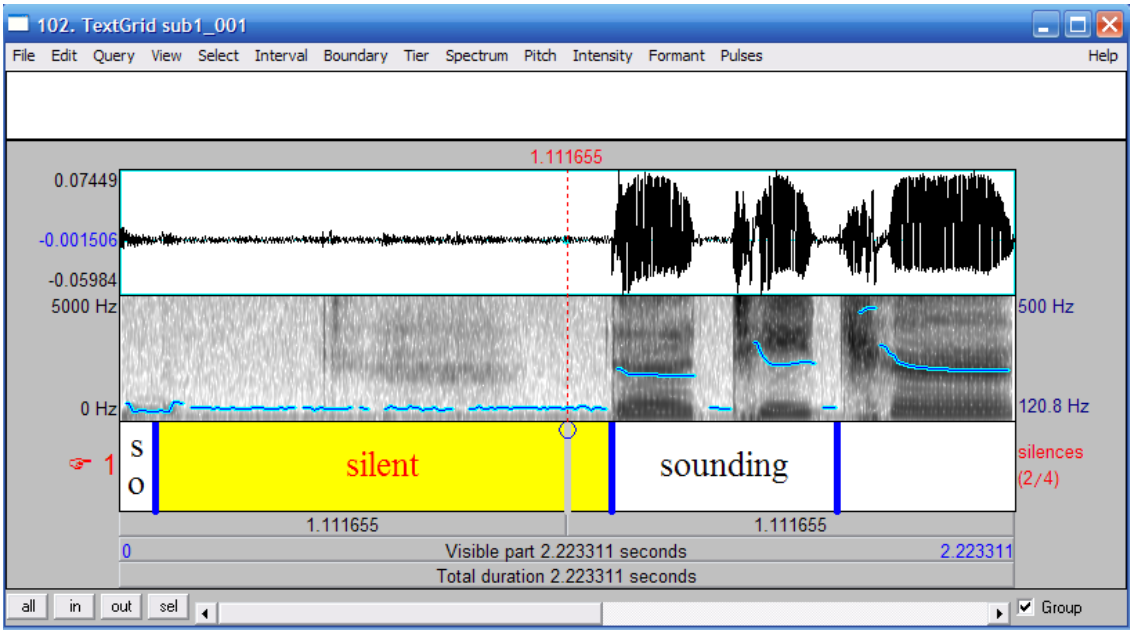


Figure 4.7: A Praat window that shows the silent interval which was used as the voice onset latency in this experiment.

4.4.2 Results

There were two timed measures, reading times, and onset latencies. The onset latency was defined as the duration between the end of sentence presentation and the onset of a connected translation. These data were screened by removing extreme data and replacing the outliers with the cut-off values which were the mean reading time or mean onset latencies plus 2.5 standard deviation from the mean. There were three proportion data: the percentage of correctly reproduced propositions in interpretation, the percentage of correctly recalled digits, and the percentage of active voice in the construction of the first clause of each sentence. Trials with false starts were not included. All of these measures were tested with ANOVA using subject (F1) and item (F2) as random factors.

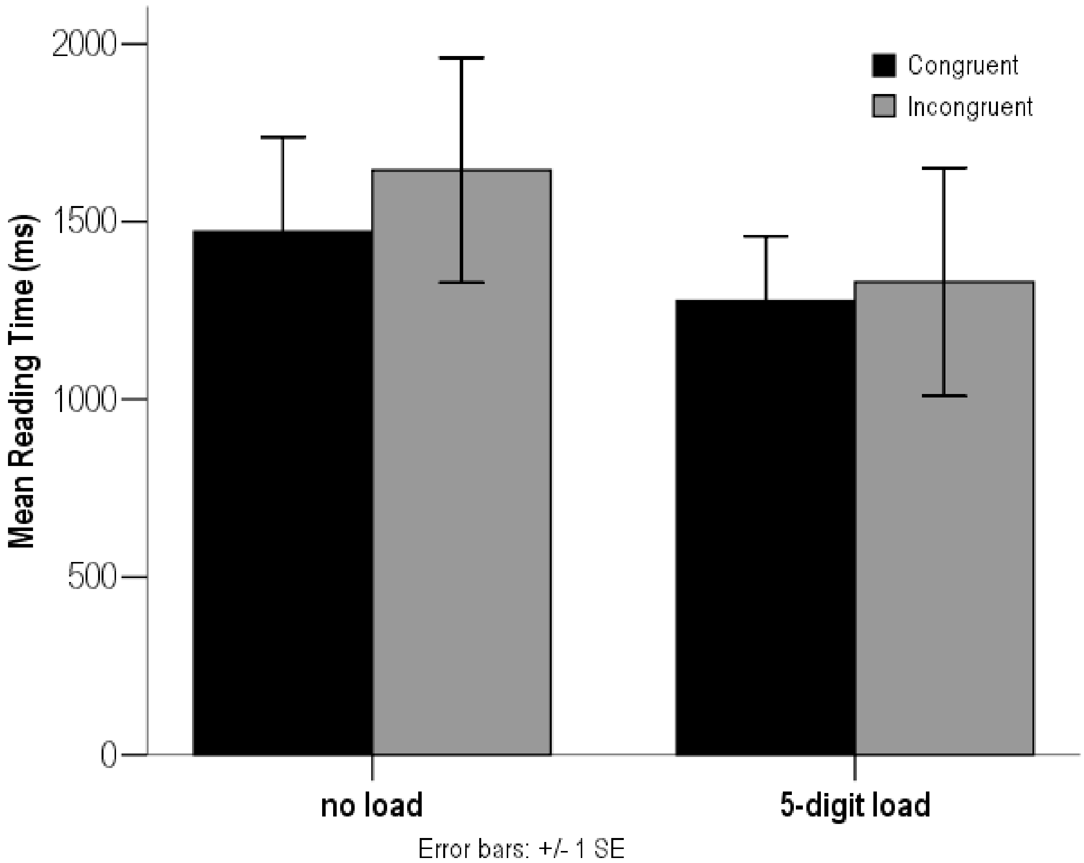


Figure 4.8: Mean Reading Time as a function of digit preload.

First, the analysis of the reading time data is presented. In by-subject analysis, the only significant effect was digit preload, $F(1,15)=4.6$, $p<.05$, indicating that reading time was reliably shorter in 5-digit-load than no-load condition (Figure 4.8, see Table A.7 for M and SD values). The same result was reported in by-item analysis in which only load effect reached significance, $F(1,47)=11.2$, $p<.01$.

As Figure 4.9 shows, onset latencies did not differ between conditions. And this was confirmed in both by-subject and by-item analyses.

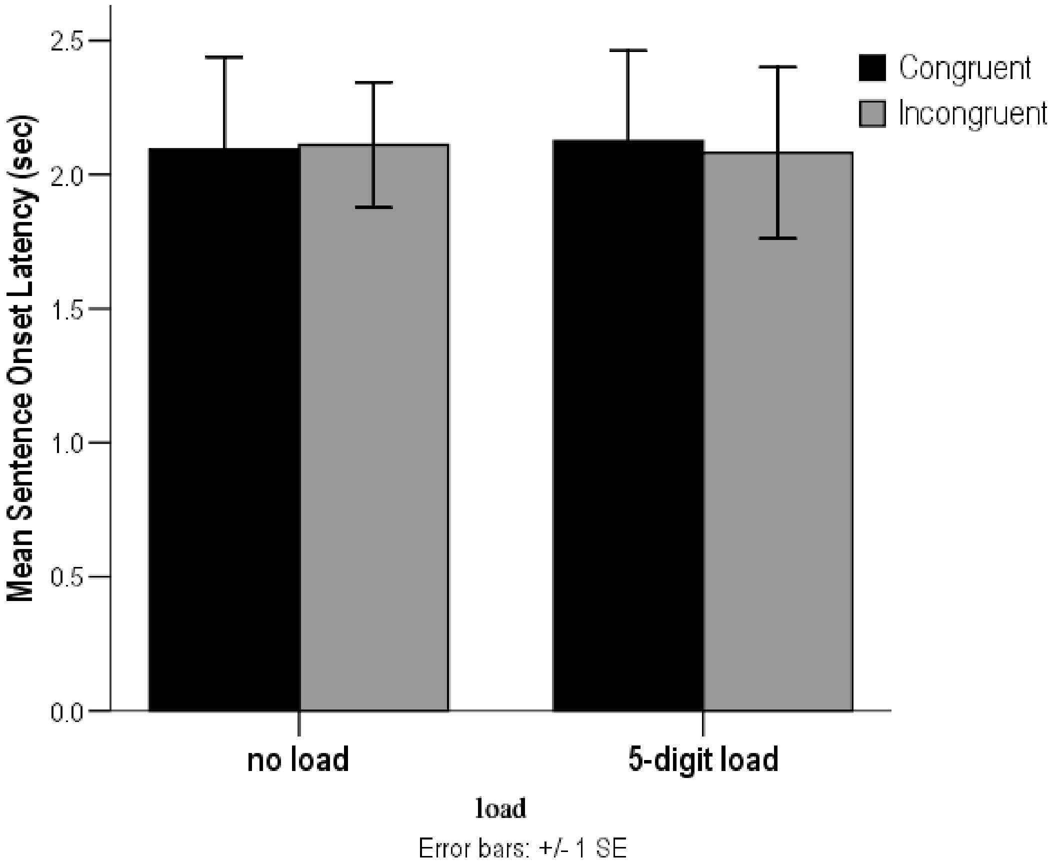


Figure 4.9: Mean onset latencies of target language production in sentence interpreting.

By-subject and by-item analyses confirmed that memory for sentences did not differ between conditions (Figure 4.10). The same was true of participants' digit recall between congruent and incongruent conditions. Post-hoc power analyses confirmed that the observed power (assuming that population effect size is equal to the sample effect size) for within-factor effect of word order congruence, digit preload, and their interaction was 0.08, 0.52, and 0.09 respectively, suggesting that future studies might increase the sample size or re-consider the working hypotheses and experimental design to afford better chance of detecting the effects of interests, e.g., congruence effect and the interaction between load and congruence.

Lastly, active voice had been more consistently used in translating congruent than incongruent sentences, $F(1,15)=16.9$, $p<.01$ (Figure 4.11). But the word order

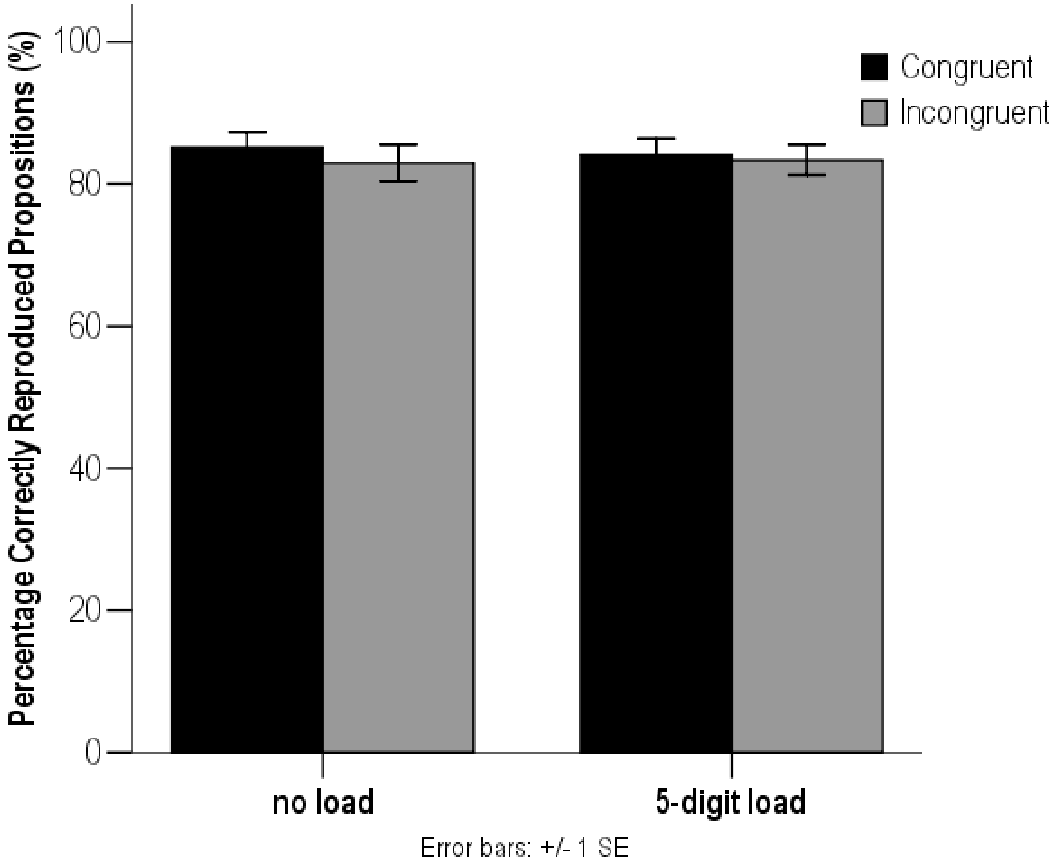


Figure 4.10: Mean percentage of correctly reproduced propositions in translation.

effect was constant in no-load and 5-digit-load conditions.

4.4.3 Discussion

One possible explanation for the unexpected results in Experiment 6 in which reading times were shorter in load conditions than when there was no load was that participants might have prioritised digit recall over reading for translation when digit recall was required immediately after sentence reading. This possibility was explored by changing the order of tasks after sentence reading: participants translated sentences into English before they recalled digits. If this manipulation encouraged them to read sentences in a way similar to that used in Experiment 5, then an interaction between digit preload and word order congruency would occur. The pattern of this interaction would have shown that reading time was much longer for

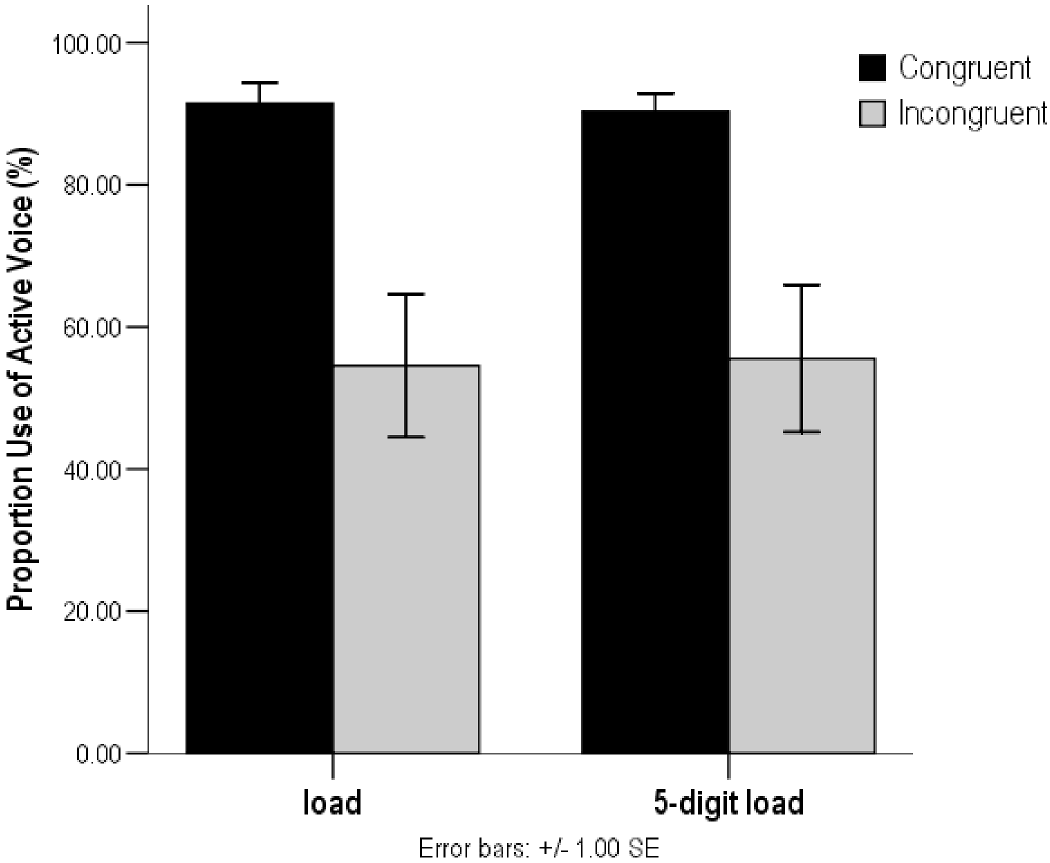


Figure 4.11: Mean percentage of active voice use in sentence construction of target language production.

incongruent sentences than congruent sentences under digit preload. In addition, reading times would have been longer in load than no-load conditions. However, the results of Experiment 7 replicated those in Experiment 6, indicating that the order of tasks after reading did not change the way participants responded to the digit preload. In line with Ruiz et al. (2008), the results in Experiments 6 and 7 therefore appear to converge on the hypothesis that TL activation during comprehension for later translation requires attentional control. However, as the alternative hypothesis relies on a null effect in word order congruence in reading times, cautions must be in order for data interpretation. This is because a null effect can be attributable to various confounds, e.g., the secondary task (digit preload) was too disruptive.

This possibility leaves the question open as to how demanding TL activation is and how the dual-task paradigm in Experiments 6 and 7 can be improved to discern the role of working memory in language interpreting. These topics are discussed in section 5.3 Methodological Considerations in the General Discussion.

The other main objective of the present experiment was to explore whether participants initiated production output faster if they had spent more time in reading sentences. Even though there was a trend of longer reading time for incongruent than congruent sentences, no differences in speech onset latencies were found in the analysis. If syntactic priming was also implicated in sentence translation, then there should have been a difference of onset latency between congruent and incongruent conditions, since the SL and TL texts share the same S-V-O structure. This prediction was not supported either. Since the data again shows that participants' memory for digits and sentence content was not sensitive to word order congruency and digit preload, the hypothesis suggested in Experiment 6 became even more plausible: participants were flexible in allocating their cognitive resource in order to optimise their performance. But there might be a premise for this argument to hold true, which is that one of the concurrent tasks has to be processed nearly automatically, otherwise there would be a trade-off between the two memory performances. The absence of the sign of parallel translation, longer reading time, seems to suggest that the resource available to participants after the digit preload took its share was not enough for them to access the TL properties. The fact that participants performed equally well with or without accessing TL properties before translation output promotes a conclusion that sentence interpreting requires working memory during comprehension only when participants have more than enough resource available, and when they consciously engage in accessing the TL properties.

One might challenge that the longer reading time as evidence for parallel translation is epiphenomenal. In other words, if participants in Ruiz et al. (2008) and this series of experiments did access TL properties during comprehension when possible, can these experiments provide evidence for parallel translation apart from reading times? In a separate analysis for Experiment 5-7, two questions were explored. 1) Did participants have an advantage in their production performance in sentence interpreting when their reading times were suggestive of parallel translation? 2) Can measures typically used in assessing speech fluency be informative regarding how working memory might be involved in sentence interpreting?

4.5 Parallel translation in sentence interpreting – Evidence from speech rate

Following the earlier argument that the null effect of word order in the first series of experiments using discourse interpreting cannot be taken to rule out the hypothesis that working memory was implicated in the stage of language production, two actions have been taken to further explore the role of working memory in sentence interpreting. First, the revised hypothesis assumed that working memory served as a workspace that interfaced SL comprehension and TL production systems. Its role in SL language comprehension has been tested by the results in Experiment 5 in which the effects of recall method and word order congruency indicated that working memory might be implicated in the activation of TL information (Macizo & Bajo, 2006; Ruiz et al., 2008). The second action was a proposal elaborated in this section, regarding how working memory might be implicated in the stage of TL production in sentence interpreting.

Recently, contrasting Levelt (1989), grammatical encoding in L1 language production has been thought non-automatic, (e.g., V. S. Ferreira & Pashler, 2002). Relative to L1 production, non-automatic processing is an even more prominent feature in second language production due to L2 speakers' limited lexical knowledge and their non-proceduralised processes in using their L2 knowledge (see section 2.1.3).

It is not surprising that among L2 speakers, including participants in these experiments, production has been found littered with pauses and fillers, which have been taken as indicators of planning (Butterworth, 1975) or problems related to lexical retrieval (H. H. Clark & Fox Tree, 2002; Maclay & Osgood, 1959). However, given chances of advance planning, speech fluency can improve. Planning has been shown to result in higher lexical and syntactic complexity (Crookes, 1989; J. Williams, 1992). Planning was also thought to alleviate the on-line demands of macroplanning and microplanning and therefore lead to faster speech initiation and more fluent speech output (Fathman, 1980; Lindsley, 1976; Lennon, 1984; Tannenbaum & Williams, 1968; Wiese, 1984). It is possbile that participants' speech rate in an interpreting task could be associated to some degree with their sentence planning. If participants' access to the information of a TL during SL comprehension is a form of advance planning, perhaps participants' speech rates would be enhanced when their TL becomes activated during sentence reading.

To date, there does not seem to be any study that directly tested the hypothesis of whether there is a relationship between planning and translation fluency. However, a series of experiments conducted by MacKay and Bowman (1969) can be potentially instructive. German-English bilinguals with German as their L1 were presented practice sentences on cards. In the practice phase, they had to repeat aloud each practice sentence at their fastest rate for twelve times. In this paradigm, when participants finished repeating each practice sentence, they would be shown a transfer sentence which they had to read aloud at their maximal speed. When they saw English sentences in practice phase, they would read out German sentences in the transfer phase, and vice versa. Experiment 1 was a norming study used to establish that there was a practice effect on speech fluency, as measured in the time taken for producing a whole sentence. A signed-ranks test comparing the mean of the first four and last four performances confirmed an effect of practice. To tease apart the source of the effect that might come from practising at the phonological, syntactic, or semantic level, two more experiments were conducted. Two types of transfer sentences were shown to participants: sentences that were either English translation of practice sentences but with a different word order or sentences that were semantically unrelated to practice sentences. Their results showed that translated sentences were read significantly faster than were unrelated sentences, suggesting that there was semantic facilitation across languages. This result confirmed one of their hypotheses in which familiarity with the meaning of a sentence increases the maximal speech rate, and the fact that the transfer sentences shared nothing but the meaning with practice sentences ruled out the account that the transfer effect could have come from practice of muscular movement in articulating one sentence repeatedly.

Another potential determinant in this transfer effect was thought by MacKay and Bowman (1969) to be the word order. Word orders very often differ between a source text and its translation, but at times, word-for-word translation produces grammatical and plausible TL sentences. To test the role of word order in the practice effect, MacKay and Bowman (1969) compared the reading time of transfer sentences that were two translation versions of practice sentences. One version was a word-for-word translation, and the other was modified in its word order (see below).

Practice sentence: Für niemand ist die Erde so viel wie für den Soldaten.

- Transfer version 1: For no one is the earth so meaningful as for a soldier.
- Transfer version 2: The earth is more meaningful to a soldier than anyone else.

In Figure 4.12, the two lines at the bottom in the right panel show that production time was clearly shorter when the translation version had the same order as

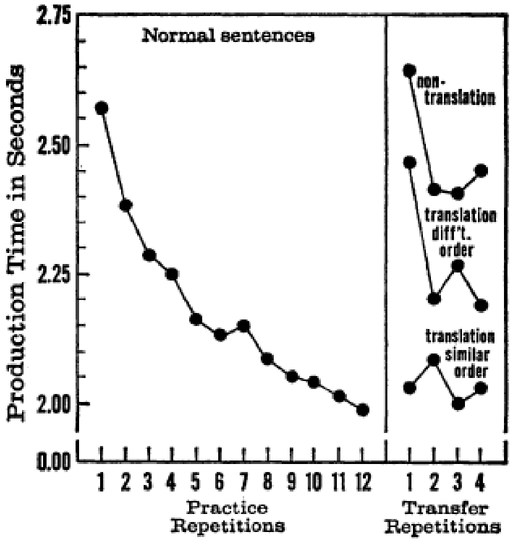


Figure 4.12: Mean duration of producing sentences during practice and transer after practice. Adapted from MacKay and Bowman (1969, p. 25).

their source sentences than when word orders differed between practice and transfer sentences. MacKay and Bowman suspected that their bilingual participants might "associate a series of words in one language with its translation in their other language," and they could "have practised the translations in the transfer phase while producing the practice sentence."(MacKay & Bowman, 1969, p. 28)

Although MacKay and Bowman (1969) took this transfer effect as evidence for the semantic facilitation hypothesis, semantic factors may have just as much role to play as shared structure in their transfer paradigm. Studies in language production have repeatedly shown that exposure to a particular syntactic structure increases the accessibility of the same structure in subsequent language processing. More readily accessible structure could reduce the processing load of generating a new sentence using the same structure (Levelt & Kelter, 1982), hence faster initiation (Corley & Scheepers, 2002; Smith & Wheeldon, 2001) and more fluent speech output (J. K. Bock & Loebell, 1990). From the perspective of structural persistence and particularly the finding of word order priming (Hartsuiker & Westenberg, 2000), MacKay and Bowman's (1969) transfer effect may be best interpreted as a joint effect of shared semantic and syntatic information between practice and transfer sentences (Loebell & Bock, 2003). A caveat was suggested by J. K. Bock and Loebell (1990) though, in which improved fluency does not result from the priming effect, rather it may arise from the repetition of the same form by the subjects. To the extent that overt planning, as in sentence repetition in the study by MacKay and Bowman (1969), can be taken as a form of advance planning for subsequent speech production, fluency measures could be a potentially informative index of the degree to which an utterance is planned ahead of articulation. In the case of sentence interpreting, it is proposed that participants' access to TL properties may be a form of advance planning, since they were informed that translation was required after sentence comprehension. The hypothesis was that if participants' longer reading time was devoted to activating TL information, it was possible that their sentence production in English would show faster speech rate in the number of syllables per second. In the following sections, the results of analysis using speech rate is presented to explore whether longer reading times were associated with more fluent speech production in sentence interpreting.

Before the speech rate data of Experiments 5-7, however, is a description of how the speech rate was calculated. Data sampling was conducted by cropping the first phrase of the translation of all experimental sentences (Smith & Wheeldon, 2001). Each segment of recording typically started with the determiner 'the' and ended with the object of the phrase, e.g. *The thief that the policeman chased* (incongruent sentence) and *The policeman chased the thief* (congruent sentence). Since the measure was the number of syllables per second, it was also ensured that each cropped recording had no silent period at either end. The phonetic analysis tool Praat (Boersma & Weenink, 2001) was used to obtain the number of syllables in each recording. De Jong and Wempe's (2009) script was used to obtain the data on two parameters: the number of syllables and the duration. Readers are referred to their paper for the description of how parameters can be set in order to identify the nuclei of each syllable. It has to be clarified, however, that Praat can identify syllables automatically but not perfectly. One reason is that the recording setting varied from participant to participant, and therefore the single parameter setting in Praat will work for some participants but not others. So it took some practice to become familiarised with optimising the Praat parameters in order to obtain more accurate identification of syllables. Because Praat was not always accurate and participants were almost never fluent, each single cropped recording has to be examined manually to obtain the accurate number of syllables. Firstly, the syllables that were misidentified or overlooked were corrected manually. Secondly, Yuan and Ellis's (2003) criterion was used to screen syllables: syllables that were filler (e.g., um), repetitions (e.g., *The thief the policeman... the thief the policeman chased*), or redundant (e.g., *The policeman that chased... the thief that the policeman chased*) were subtracted from the total number of syllables. The result of this two-step screening is the accurate number of productive syllables (Figure 4.13).

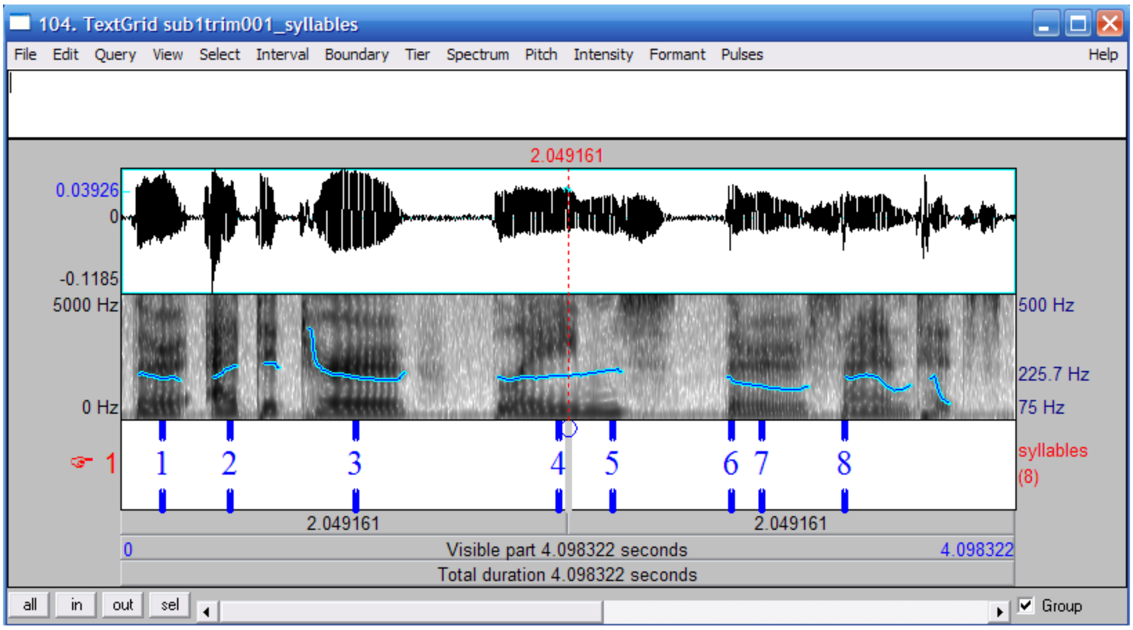


Figure 4.13: Speech rate as the duration of the first clause of each sentence divided by the number of productive syllables. The syllables are the nuclei captured by Praat.

The measure of speech rate on the basis of the productive syllables can be thought of as a measure of meaningful productivity (Yuan & Ellis, 2003). However, not every trial was included in the speech rate analysis. Participants did not always translate the first clause in full, and even if they did, they showed problems in lexical retrieval which took two forms. Sometimes participants use generic term or hypernyms – they used *the guy* for the word *accountant*, or *leader* to replace *principal*. Other times, they described and defined the word that they could not retrieve, e.g. *The guy who played tricks to make people laugh* for the word *clown*. These phenomena have been documented in studies of communication strategies among L2 speakers (Dornyei & Kormos, 1998), but trials like these were excluded from the analysis as these strategies might indicate that these words were either absent in their lexicon or very infrequently used. The procedure described above was applied to Experiment 5-7. The following sections report the results for each experiment in turn.

4.5.1 Speech rate analysis of Experiment 5

Since the focus of the analysis was the speech rate of translation, the analysis of Experiment 5 excluded the repetition conditions. So the dependent variable was the speech rate, and the independent variable was the word order congruency.

A repeated measure ANOVA showed that participants' speech rate did not differ between congruent (2.5 syllables/s) and incongruent (2.6 syllables/s) conditions in by-subject ($F(1,15)=0.48$, $p=.51$) or in by-item analysis ($F(1,47)=0.739$, $p=.394$). So even though participants did spend more time in reading words in translating incongruent than congruent sentences, they did not show an advantage of the potentially early access to the TL lexical and syntactic information in their speech

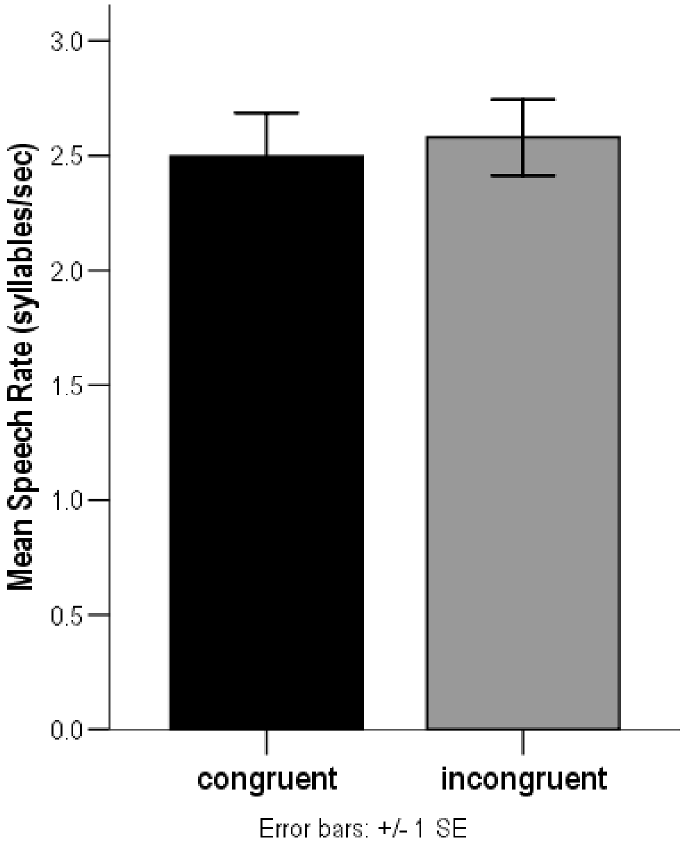


Figure 4.14: The speech rate in congruent and incongruent conditions of sentence interpreting.

production. As it was very difficult to obtain the data of the speech onset in Experiment 5, it remains an unanswered question as to whether there was an advantage of speech onset in translating incongruent over congruent sentences.

4.5.2 Speech rate analysis of Experiment 6

Experiment 6 manipulated digit preload (no load vs. 5-digit vs. 7-digit) and word order (congruent vs. incongruent) in a repeated-measures design. Again, a repeated measure ANOVA was used to compare participants' speech rates between conditions.

Figure 4.15 shows that there did not seem to be a load effect, but there was a trend of higher rate for incongruent than congruent sentences (see Table A.8 in the

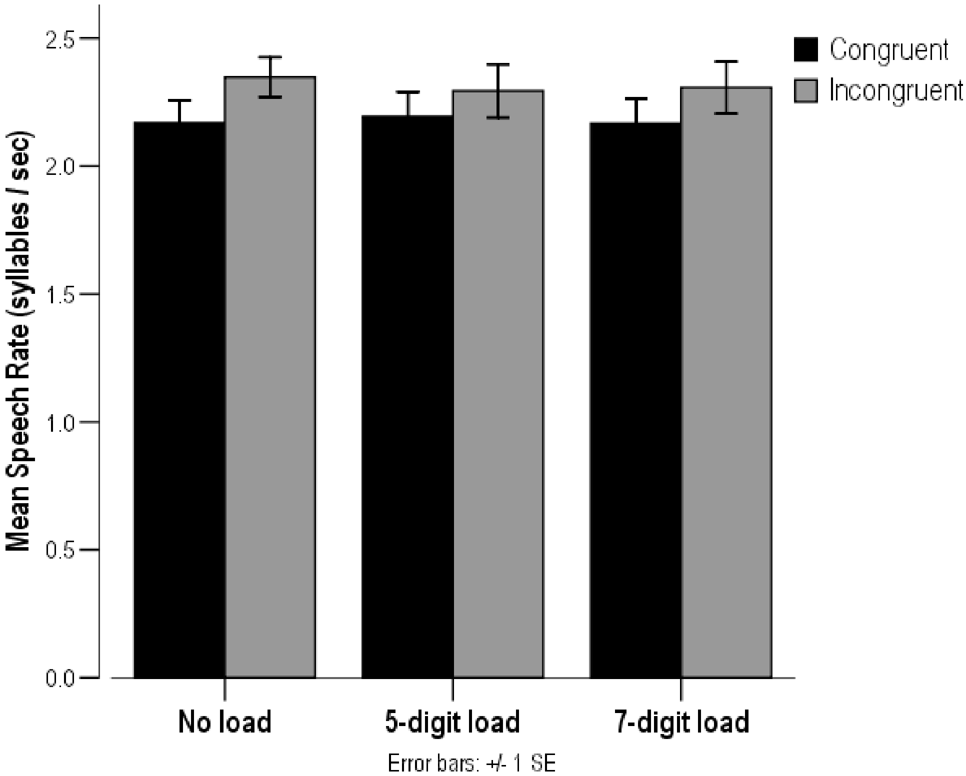


Figure 4.15: The speech rate as a function of word order congruency in sentence interpreting.

Appendix for *M* and *SD* values). The ANOVA confirmed the word order effect in by-subject analysis, $F(1,17) = 6.3$, $p < .01$. But this effect was not significant in by-item analysis, $F(1,47) = 2.4$, $p = 0.13$. This result indicated that participants might have benefited from slightly longer reading times in incongruent than congruent conditions, even though the word order effect was only significant in the no-load condition in Experiment 6. Although the differences of reading times in load conditions were not reliable, participants' were slower numerically in reading incongruent than congruent sentences. Thus the speech rate analysis indicated that resources invested in interfacing SL comprehension and TL production system prior to sentence interpreting might have paid off in the stage of TL production since processing load may have been reduced, because the lexical and syntactic information may still be highly activated and accessible in participants' working memory.

To the extent that the advantage of early access to TL information was reliable, a very interesting question would be what happens when TL production was under digit load? Recall that Experiment 6 and 7 differed in the order of tasks after sentence reading: participants translated before recalling digits in Experiment 7. This procedure implied that while participants were translating a sentence, they might somehow have to hold the digits throughout the duration of TL production. In other words, there appeared a dual-task situation in the 5-digit load condition in Experiment 7 while the TL production in Experiment 6 was load-free. If working memory was implicated in holding the activated TL information, then the requirement to buffer the preloaded digits could leave little capacity to the former function and therefore the advantage of early access to the TL may be removed.

4.5.3 Speech rate analysis of Experiment 7

It was predicted that if working memory was required for holding the TL information activated prior to the stage of production, the requirement of digit recall after translation might demand too much working memory to leave enough capacity for the early TL access to show its advantage during sentence production in English. Experiment 7 crossed factors of load (no load vs. 5-digit load) with word order (congruent vs. incongruent) in a repeated measures design.

Figure 4.16 shows that speech rate did not appear to differ in no-load condition, but there was a trend of higher rate in incongruent than congruent condition in the load condition. An ANOVA however showed that there was no main effect of word order congruency in both by-subject, $F(1,15)=.05$, $p=.83$, and by-item analyses, $F(1,47)=.54$, $p=.47$. As there was no interaction between load and congruency, the results did not allow a conclusion that the absence of advantage of early TL activation was attributable to the working memory demand for holding the digits.

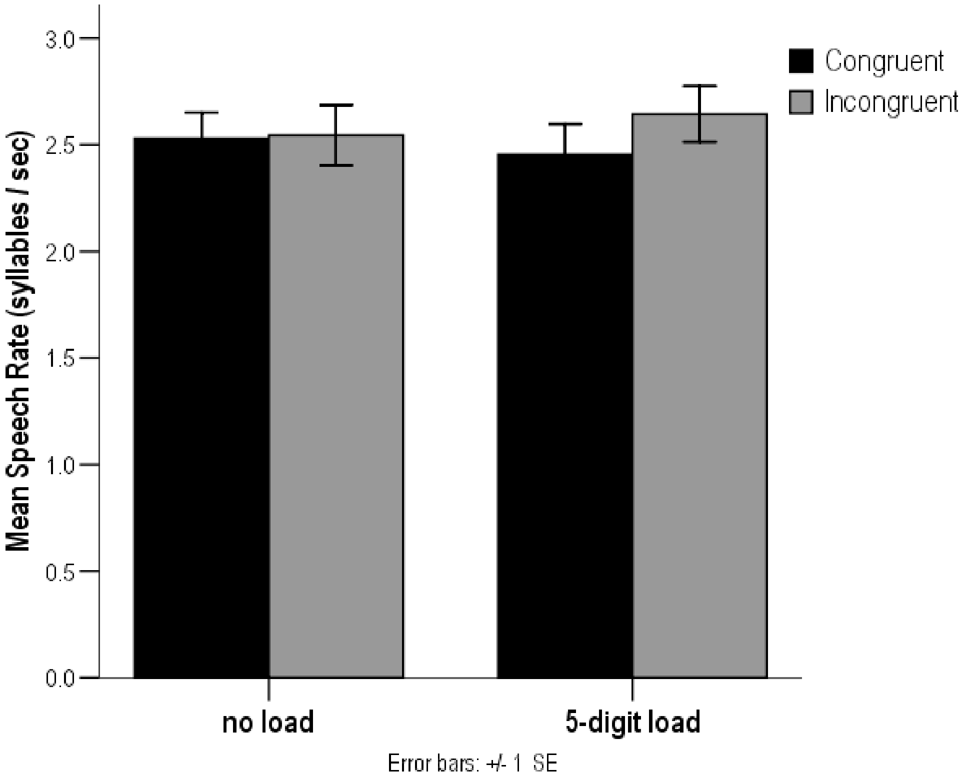


Figure 4.16: The speech rate as a function of word order congruency in sentence interpreting.

As the reading times in Experiment 7 showed no word order effect even in the no-load condition, together with the absence of difference in speech onset and speech rate between congruent and incongruent sentence interpreting, it was possible that participants in Experiment 7 carried out the reading task differently from those in Experiment 5 and 6. One possible explanation could be their education or language background. Although all participants were native Chinese speakers, those in Experiment 5 and 6 came from Taiwan, and those in Experiment 7 were from the People's Republic of China. One feedback from a participant in Experiment 7 was that she found the wording of Chinese object relative clause a bit awkward. According to her, it would sound more natural if the clause 警察追趕的小偷 (the thief that the policeman chased) has a particle 被(bei) at the sentence initial position: 被 (bei)警察追趕的小偷 (the thief that was chased by the policeman). The particle 'bei' is typically used in passive sentence construction in Chinese.

被	警察	追趕	的	小偷
bei	policeman	chased	DE	thief

It has been suggested that 'bei' construction is used to signal adversity, e.g., 我的手錶被(bei)偷了 (My watch was stolen). However there are increasingly more and more nonadversity usages of the bei construction in modern Chinese. This development was thought to be a result of the influence of Indo-European languages, especially English (C. N. Li & Thompson, 1981). The comment made by the participant about the usage of bei in object relative clause in Mandarin Chinese was valid that some object relative clauses sound more natural when they are in passive voice. But not all verbs that participate in object relative clause construction allow alternation between active and passive voices. For instance, the verb *contact* in *the major who the accountant contacted* does not allow a passive construction in Chinese. There are a few such verbs that appeared in the testing material such as *know*, *marry*, *avoid*, and *visit*. Recall the usage of passive and active voices in Experiment 6 and 7 in which the percentage of active usage did not differ between different load conditions. Originally, it was suspected that the motivation of using passive construction was to avoid perspective shift (MacWhinney, 1999) from the object to the subject, i.e., participants did not shift perspective when they used passive voice, but they had to shift their perspective from the object 'thief' to the subject 'policeman' if they decided to use active voice in constructing *the thief that the policeman chased*. Since there was no effect of preload on the frequency of passive construction in objective relative clauses, the possibility of using passive voice as a device to reduce cognitive load was weakened. It appears more plausible that the reason for participants' less consistent use of one voice or the other in object relative clause construction may be a frequency-based decision: some Chinese verbs might show higher frequency in passive relative clause construction than others. However, since all to-be-translated Chinese sentences that contain object relative clauses used active voice, the choice between active or passive voice in producing

English translation would seem to depend on participants' reading. If a sentence received a passive reading, then its English translation would use passive construction, or vice versa. Given that the comment made by the participant can generalise to other participants, it was possible that experimental sentences had received active or passive reading and were encoded in their memory before translation began. If this account holds, then it can be taken as supportive evidence that highlights the role of both comprehension and memory components of translation. Regarding the comprehension component, the way that participants might have given a phrase using active construction a passive reading supported the CI model (Kintsch & Van Dijk, 1978) in which comprehenders normalise a discourse with devices such as reordering, lexical substitutions, or perspective changes (also see Luftig, 1982). In addition, because this particular phrase has been encoded with a notion of passive voice, the representation in their working memory promoted the use of passive construction in TL production.

4.5.4 Summary of Speech Rate Analysis

Motivated by exploring the role that working memory might play in language interpreting, the self-paced reading paradigm used by Ruiz et al. (2008) was adopted and the word order difference in noun phrase modifier between Chinese and English was used in designing testing material. Reading times data in Experiments 5 and 6 replicated the finding of Ruiz et al. (2008) that participants were slower when they read incongruent sentences than congruent sentences for later translation. Unexpectedly, reading times were shorter in digit preload conditions than no load conditions in Experiment 6 and 7, perhaps because the resource demand for maintaining the digit memory overloaded participants' working memory, thus they had to give up parallel translation that appeared to be implicated in no-load condition. Apart from the resource limitation, faster reading time might have also been driven by participants' strategy to reduce the time lapse between digit presentation and recall, i.e. the retention interval, so that the chance that digit memory decayed may be reduced.

Unlike studies such as Fedorenko et al. (2007) in which verbal-mediated tasks interacted, the unexpected pattern of faster reading in load than no-load conditions and the absence of interaction between load and word order congruency in Experiment 6 and 7 made data interpretation difficult using a shared-resource account. In addition, unlike Waters et al. (2003) who found that large digit load resulted in longer reading times and poorer digit recall, participants in Experiments 6 and 7 showed no memory decrement in sentence content or digit memory in load conditions. Since the concurrent load did not compromise the construction of sufficient representation of a sentence for later translation and there was no trade-off between the two tasks, working memory that was used to process Chinese sentences for later translation might be minimal. Even though participants in Experiment 6 did not seem to apply parallel translation in load conditions, the pattern of numerically longer reading time in incongruent than congruent conditions was in the predicted direction. Most importantly, their speech rates were significantly higher in translating incongruent than congruent sentences in both no-load and load conditions, suggesting that resources invested prior to TL production were rewarded with faster encoding in their L2, English.

4.6 Experiment 8 – A replication of Experiment 5 using Chinese Ba construction

Although Experiment 5 and 6 replicated the findings of Ruiz et al. (2008) that were indicative of parallel translation, one limitation in the generalisability of the results stemmed from the material, in which the congruent and incongruent versions of experimental sentences differed slightly in their meanings. For instance, the first clause of a congruent sentence could be 警察追趕小偷 (*the policeman(警察) chased(追*

趕) the thief(小偷)), while its corresponding incongruent version was 警察追趕的小偷 (*the thief(小偷) that(的) the policeman(警察) chased(追趕)*). Even though the two versions only differed in the addition of Chinese particle 'DE' in incongruent sentences, it was not clear whether reading time differences between the two versions could be partially due to the garden-path effect in processing a relative clause[1]. In order to replicate Ruiz et al. (2008) with a more stringent design, the experimental procedures were kept the same as used in Experiment 5, but a whole new set of sentences which used the Chinese Ba construction were generated. Structurally, the Ba construction is straightforward. With the introduction of coverb 'Ba', the S-V-O word order is converted into S-Ba-O-V without considerably changing its meaning (see the phrase structures and example sentences below). An English transitive verb phrase using telic verb can be translated into a Chinese Ba VP or a non-Ba VP. Therefore, when translating a Chinese Ba construction into English, their word orders are incongruent. But when translating a Chinese non-Ba transitive VP into English, their word orders are congruent.

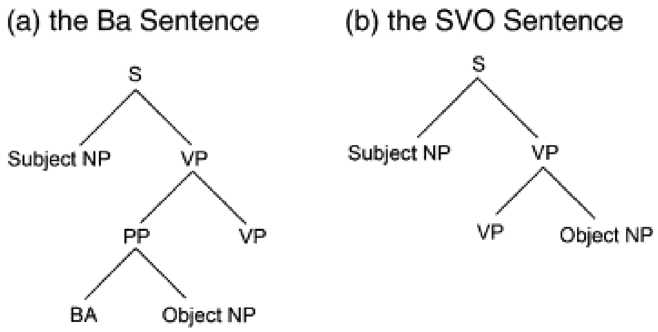


Figure 4.17: Phrasal structures of Chinese Ba and non-Ba sentences, adapted from Ye (2007, p. 137).

湯姆把(Ba)花瓶打破了然後要我買個新的

30. **Non-Ba version**:

Tom(湯姆) <u>broke</u>(打破) <u>the vase</u>(花瓶) and(然後) asked me(要我) to buy a new

[1]This comment was provided by Charles Clifton in the author's presentation (Jin, 2008) in the postgraduate psycholinguistic conference, held in the Department of Psychology, University of Edinburgh in 2008.

one(買個新的).

31. **Ba version**:

Tom(湯姆) <u>BA</u>(把) <u>the vase</u>(花瓶) <u>broke</u>(打破) and(然後) asked me(要我) to buy a new one(買個新的).

Although it seems a very straightforward manipulation of word order, not all Chinese verbs permit Ba construction. Sentence 33 is not acceptable:

她把(Ba)那隻小貓愛

32. **Non-Ba version**: She(她) loves(愛) the kitten(那隻小貓).

33. **Ba version**: She(她) Ba(把) the kitten(那隻小貓) loves(愛).

The reason that the verb *love* cannot participate in Ba construction is that it violates the rule that a Ba sentence usually involves telicity: "something happening to the entity referred to by the Ba noun phrase" (C. N. Li & Thompson, 1981). The verb *love* does not indicate an action that has an end state or an action being complete like the verb *break* does. The consequence of breaking a vase is clear, but there is no specific consequence of a kitten being loved. Since it was inappropriate to switch non-Ba to Ba-sentences arbitrarily, and the author was not aware of any study that documented a list of Chinese verbs that can participate in Ba construction, the Academia Sinica Balanced Corpus of Modern Chinese was used in material preparation.

As the present experiment was designed to replicate the findings in Experiment 5, it was hypothesised that the word order manipulation that resulted in incongruent condition between Chinese Ba sentences and their English translation would lead to longer reading times than reading non-Ba Chinese sentences for later translation. Similar to Experiment 5, it was also predicted that there would be an effect of recall method: reading times would be longer in reading for translation than reading for repetition conditions.

4.6.1 Method

Participants

Sixteen native Chinese speakers (5 males, 11 females) from Taiwan were recruited in the University of Edinburgh in this paid experiment (8 pounds each). They were aged between 22 and 41 (M= 26.7). They had IELTS scores of 6.5 or higher (M= 6.9, SD=0.4). One of them took part in Experiment 5. However, Experiment 5 and 8 used different materials, therefore, this participant's data was not excluded from the analysis.

Procedure

As this experiment replicated the procedure of Experiment 5, participants read sentences without digit preload for later repetition or translation.

Material

Forty-eight Ba sentences that used different Chinese verbs were selected from the Academia Sinica Balanced Corpus of Modern Chinese. These selected sentences were then modified in such a way that each conjoint sentence was composed of a Ba-phrase, followed by a phrase that described the consequence of the event in the Ba-phrase. These sentences were subject to independent rating on their syntactic naturalness and plausibility in a web-based survey. Since the experimental sentences were rated in the same survey as described in Experiment 5, the procedure is not repeated here. t-tests confirmed that there was no difference in the naturalness rating, $t(47)$ = 1.43, p =.16, between BA (M = 5.0, SD =0.65) and non-BA sentence (M=5.1, SD =0.63, nor was there a difference in plausibility, $t(47)$=1.38, p =.17) between BA (M = 5.1, SD =.68) and non-BA sentences (M = 5.2, SD =.62). The result indicated that Ba and non-Ba versions are equated in their naturalness and plausibility.

These experimental sentences were separated by 48 fillers and presented in a pseudo-randomised order in two blocks of 96 trials. By using a Latin Square design, four lists of sentences were generated so that participants would read either Ba- or Non-Ba version of each experimental sentence, but both versions of each sentence were read for repetition or translation. The experiment was blocked by recall method. Half of the participants started off with reading-for-translation block, followed by reading-for-repetition and the other half had the opposite block order.

4.6.2 Results

As this experiment was designed to explore whether Chinese Ba construction that resulted in word order difference between Chinese and English would replicate the findings in Experiment 5 in terms of reading times, only reading times were subjected to statistical test. Unlike the analysis in Experiment 5 which used the reading time of one critical word, it was not possible to use the same procedure in the present experiment. This is because the word order in Ba sentences was different from that in non-Ba sentences:

Non-Ba sentence: Tom <u>broke the vase</u> ...
Ba sentence: Tom Ba <u>the vase broke</u> ...

Since it was inappropriate to compare the reading times of *vase* and *broke*, it was decided to compare the sum of reading times of the verb and its object, i.e. the underlined area in the example above.

Figure 4.18 shows that reading times were longer in reading for translation than reading for repetition (see Table A.9 in Appendix A for condition M and SD values). A repeated ANOVA confirmed the effect of recall method in by-subject, $F(1,15)= 24.2$, $p<.05$, and by-item analysis, $F(1,23)= 150.1$, $p<.01$, but there was no effect of word order congruency. Critically, the two factors did not interact.

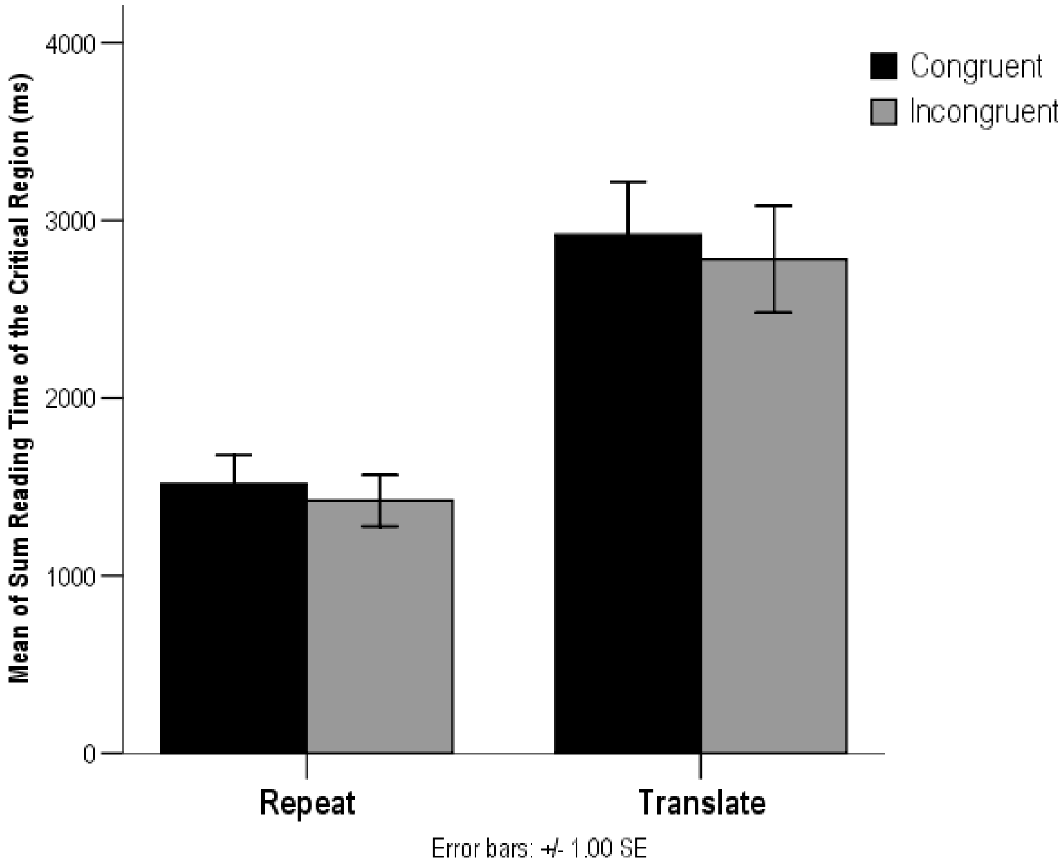


Figure 4.18: Reading times as a function of recall method in sentence interpreting.

4.6.3 Discussion

In this exploratory study, a new set of sentences were used that could meet the requirement of word order manipulation and also avoid the criticism that relative clauses attracted in Experiment 5-7. The results replicated Experiment 5 only in one aspect, the effect of recall method. Surprisingly, there was even a trend of faster reading for incongruent sentences in both reading-for- repetition and reading-for-translation conditions, contrasting the longer reading time for incongruent sentences in Experiment 5. There was in fact a study that reported faster reading for Zoeng-construction (Cantonese equivalent of Ba-constructions in Mandarin Chinese) compared to that of double-object phrases. According to Cheung (2007), the faster reading times and higher accuracy in imitating Zoeng-construction indicated that Zoeng-construction might have been preferred by readers as it required less resource in processing and production. Because the comparison of reading times in the present experiment was between Ba-construction and simple dative construction, rather than double object dative construction, Cheung's (2007) account does not readily generalise to the result of this experiment. It was possible that the effect of word order might be modulated by the degree to which structures differed between an SL sentence and its TL translation. The way with which this difference can be formalised cannot be decided yet. Intuitively, the distance between the constituents that change positions as a result of translation could be one predictor of how likely the word order effect might appear. If this were a plausible account, then one would expect to find word order effect in translating sentences that resulted in a large word order difference, e.g., Experiment 5, but not in cases where word order difference as a result of translation was relatively smaller.

The above hypothesis, however, cannot explain why a word order effect was evident in the study by Ruiz et al. (2008) in which the structural difference was a result of swapping between immediately adjacent NP and its adjective modifier in Spanish-English translation, e.g. nice house vs. casa bonita (house nice). So the question is why similar word order difference in terms of the distance between constituents that take different positions in a TL led to a difference in reading time in Ruiz et al., 2008 but not in the present study? There is at least one possibility that warrants further exploration, and that is the word order effect on reading time in a reading-for-translation task might be more sensitive to the type of phrase structure than the physical distance between constituents. To avoid arriving at a premature conclusion that NP modifications (e.g., *nice house* in Ruiz et al. (2008) or *the policeman that chased the thief* in Experiment 5 of this thesis) are more likely to be translated using a parallel translation strategy than VP modification(e.g., *Ba window broke* in the present experiment), future research is required to replicate the paradigm by permitting physical distance of involved constituents and the types of phrase structures to be orthogonally manipulated.

4.7 Summary

To explore the role that working memory might play in language interpreting, the second series of experiments adapted the paradigm used by Ruiz et al. (2008) that echoed proposals for interpreting research made by De Groot (1997) and P. Padilla et al. (1999). De Groot proposed that "one approach to take in the study of translation is to focus on stimuli smaller than a complete text, for instance, words and sentences. The goal would be to discover the factors that affect translation of that particular language unit. This approach departs from the assumption that translating a complete text must depend at least partly on the ease with which the elements that constitute the text are translated when presented in isolation"(De Groot, 1997, p. 32). The use of sentences as testing material in this series of experiments made material control easier and results more tractable than using complex and long discourses. P. Padilla, Bajo, and F. Padilla suggested that a dual-task paradigm can be used to study "the impact of joint performance and therefore identify possible interactions and modifications."(P. Padilla et al., 1999, p. 73). Although P. Padilla et al.'s (1999) proposal was motivated by the study of simultaneous interpretation, Experiments 5-7 of this thesis have shown significant usefulness of the dual-task paradigm in exploring the potential role working memory played in sentence interpreting in the form similar to consecutive interpreting. Experiments 5-7 also demonstrated that the use of both on-line measures (e.g., reading times, speech rate) and off-line measures (e.g., memory performance for sentences and digits) could complement each other in studying a research topic that engaged different languages and both comprehension and production systems. The failure of Experiment 8 in replicating the findings in Experiment 5 must not be seen as a straightforward falsification of parallel translation. Instead, it served to provoke interdisciplinary discussions and stimulate future research in order to elucidate the processes in language interpreting.

Chapter 5

General Discussion

Language interpreting is a difficult task for inexperienced interpreters, but it is also a complex subject for researchers. In two series of experiments, this thesis attempted to address 1) whether or not working memory is involved in language interpreting; 2) if working memory is implicated in language interpreting, how? In this chapter, the findings from these experiments and their implications will be discussed.

While the research questions that this thesis set out to explore primarily concern the role of working memory in language interpreting, some empirical evidence suggests that the stage where working memory might be implicated appears to be contingent on the assumption of when language recoding occurs during language interpreting (Macizo & Bajo, 2006). If language recoding occurs during SL comprehension, and it incurs a cost, e.g., longer processing, then it could imply that working memory is involved in language recoding. *Recoding* or *transformation* is an important hypothetical stage or process in language interpreting, whereby the meaning of the SL discourse is encoded in an interpreter's TL (Christoffels & De Groot, 2004). Accoring to Christoffels and De Groot (2004), transformation might have subcomponents of reformulation and language switching. However, little is known about its mechanism or temporal course during language interpreting. For instance, it could be asked as to whether recoding takes place during SL comprehension or during TL production, or indeed during both processes. Three hypotheses have been made for how recoding might occur between a SL and its TL in an interpreting task. One is sequential translation which is based on the meaning of a whole (or partial) sentence; another is parallel translation which is based on the form; and the third one is a hybrid between the previous two (Christoffels & De Groot, 2005; De Groot, 1997). By assuming sequential translation in discourse interpreting, where language recoding is assumed to take place at the stage of TL production, Experiments 1 to 4 explored the role of working memory in language interpreting.

If working memory is concurrently implicated both in grammatical encoding into the target language, and in temporary storage of the discourse content, then higher demand in one function might compromise the other. Thus discourses that differ in word order between languages could increase the processing load and leave less resource for memory maintenance, affecting recall performance. In Experiment 1, Chinese-English bilingual participants' memory performance was compared when they translated passages from Chinese to English and from English to Chinese, where the expected word order was either congruent or incongruent between source and target. Recall was not sensitive to word order or direction of translation. Perhaps surprisingly, memory for incongruent discourses was numerically better than that for congruent sentences. Three follow-up studies were conducted to explore the unexpected results. Designed to explore the impact of the differences of general language proficiency and training between proficient bilinguals in Experiment 1 and students that were trained to be interpreters, Experiment 2 revealed that interpreting trainees performed just like the participants in Experiment 1 did, suggesting that memory performance was not modulated by translation direction in proficient translators. Experiment 3 explored the relationship between surface form transformation and recall. As discourse paraphrasing did not result in better recall than verbatim recall, it was concluded that the better memory performance for incongruent discourse interpreting suggested by Experiment 1 was not the result of active manipulation of word form or word order in interpreting. Finally, a free recall task among native English speakers in Experiment 4 showed that the incongruent discourses tested in earlier experiments were intrinsically more memorable than congruent discourses. Despite this confound, this series of experiments highlighted the importance of comprehension in interpreting. It also showed controlling for various factors, e.g., material design, in empirical studies of language interpreting is a challenge facing researchers.

In order to address the research questions and issues related to experimental design in Experiments 1-4, Experiments 5-8 adopted the procedure of Ruiz et al. (2008) and introduced the dual-task technique in Experiments 6 and 7. The original hypothesis that assumed a strict sequential translation in Experiments 1-4 was modified in Experiments 5-8 to accommodate that, 1) parallel translation might occur, i.e., recoding or transformation might take place during the comprehension phase of interpreting; and 2) working memory could be implicated in both comprehension and production phases of sentence interpreting. Experiment 5 replicated a self-paced reading study by Ruiz et al. (2008), comparing participants' times to read sentences for translation to those for reading them normally. The data showed that participants accessed lexical and syntactic properties of a target language in the reading-for-translation condition when resources were available to them. The interaction between recall method and sentence congruency could indicate that extra resource was required for TL activation during SL comprehension (Ruiz et al., 2008). However, in Experiment 6, when participants' working memory was occupied by a secondary task (digit preload), reading times were only different numerically between congruent and incongruent sentences. Surprisingly, contrasting the predictions, reading times decreased as digit preload increased and there was no interaction. Experiment 7 refined the procedure in the order of responses for the dual tasks but replicated the results of Experiment 6. Finally, another set of sentences that used Chinese Ba construction were used in Experiment 8 in an attempt to replicate Experiment 5. A failure to replicate the earlier findings suggested that working memory demand might differ for different syntactic structures in sentence interpreting.

A few findings warrant further discussion. The first section debates the role of working memory in language interpreting. In a following section, the psychological reality of parallel translation is discussed. And the final section concerns the methodologies used particularly in Experiments 5-8. In each section, directions for future research are also suggested.

5.1 Working memory in language interpreting

By extending Waters et al.'s (2003) dual-task approach, it was predicted that concurrent load would increase processing time in the self-paced reading task. The unexpected results beg a question as to whether dual-task is suitable in addressing the research question. Conventionally, in a dual-task paradigm that studies syntactic processing, reading is the primary task while the choice of the secondary task depends on a particular research question. When it is applied to the reading-for-translation task, there could actually be more than two tasks that are performed by participants. In addition to reading, and the secondary task (e.g., digit preload), the third is the process in focus – parallel translation. Unless it can be assumed that reading itself does not demand resource, it would be difficult to interpret whether the result was driven mainly by the external load. But given that there was no reading time difference between reading congruent and incongruent sentences for later repetition, and the effect of recall method in Experiment 5 was rather large, the issue of multi-tasking may not be a major concern. Before a better paradigm becomes available, it is proposed that dual-task methodology can still be used in exploring how working memory is involved in language interpreting, but the paradigm in this thesis can be improved in some respects.

The reason that the reading time data were not in the predicted direction may be that the digit preload was too demanding. Hegarty, Shah, and Miyake (2000) discussed two factors that can limit the applicability of the dual-task paradigm. One was the existence of a central bottleneck for response selection. Of direct relevance to the present thesis was the strategic trade-off between primary and secondary task performance. Some studies have reported that participants tend to allocate more resources to the task that they perceive to be more demanding (e.g., Bourke, Duncan, & Nimmo-Smith, 1996; Navon & Gopher, 1979). In the dual-task paradigm used in the present thesis, digit preload might be more demanding than parallel translation. Therefore, participants might have allocated more resource to memorising digits, with little available for use in parallel translation. In order to avoid overloading participants, future research could use smaller load, e.g., a three digit preload.

Regarding production data, while reading times decreased significantly in load conditions in Experiment 6 and 7, there was no performance decrement in recall for sentences or digits. Contrasting the original assumption regarding WM as a unified capacity, these results were more consistent with the SSIR theory in which the resource used in remembering digits was separate from that supporting reading comprehension. Relevant to the first point, WM might not be shared by two or more tasks. Instead, the results might fit the switching account better: "trade-off may take place only when working memory is not fully occupied by processing"(Towse & Hitch, 1995, p. 123). It has been suggested that participants appeared to switch between operations when they were preloaded with six digits in a sentence verification task (Hitch & Baddeley, 1976). The participants in Experiments 6 and 7 of this thesis might have realised that the concurrent tasks had barely left them with any resource for parallel translation to be carried out. In order to optimise their performance, participants might have switched resource between memorising digits with subvocal rehearsal and sentence reading. As there was no overt control over participants' rehearsal, the explanation for their faster reading time under load is speculative at the moment. In the memory decay hypothesis, one way of preventing the strength of memory traces from decaying is to reduce the time lapse during which information has to be retained (Barrouillet, Bernardin, & Camos, 2004). The time-based resource-sharing (TBRS) model (Barrouillet et al., 2004) perhaps could throw some light here. In the TBRS model, it is assumed that "an attentional mechanism rapidly alternates between refreshing memory and performing the distractor task. Thus, forgetting is a function not of absolute time but of the proportion of time that is occupied by the distractor task (cognitive load)"(Lewandowsky, Oberauer, & Brown, 2009, p. 123). Under digit preload, in order to prevent the digits from decaying, participants might attempt to reduce the proportion of time of the digit retention interval that was occupied by the 'distractor task', which was the reading task, leading to shorter reading times compared to those in no-load conditions. This line of reasoning should support the proposal made earlier in using a smaller digit preload in future studies if the same paradigm is to be used. In doing so, the load may be less likely to fully occupy participants' working memory, allowing the possibility of time-based resource-sharing processing in conducting concurrent tasks.

While the time-based resource-sharing account remains to be verified, there is another possibility for further exploration. Another reason that Experiments 6 and 7 failed to provide supporting evidence that implicated working memory by using digit preload in the dual-task paradigm was that parallel translation might be supported by another component of working memory rather than the phonological system. In addition to exploring the role of the phonological system using smaller preload, another candidate component is the CE. As mentioned earlier, language interpreting was thought to require participants' both languages to be highly active, if not equally so (D. Green, 1993). For less proficient bilinguals, to switch from a highly activated language that is currently used in reading to the other language in order to access translation equivalents could entail a switch cost (Seidenberg & McClelland, 1989). The asymmetry of switch cost in picture naming that requires participants to alternate between languages suggests that more inhibition is needed to suppress L1 when production in L2 is required, but the consequence is that when L1 is required in the next trial, it takes longer to be re-activated. In the context of parallel translation during reading, in order to access the translation equivalent in TL, readers might need to temporarily suppress the SL. Michael et al. (2002) suggested that working memory could be a key player in the role of bilingual control (see paragraph **Bilingual Memory: What is activated in lexical access?**). The suppression account in switching between languages finds some supporting evidence from neuroimaging studies. In a word translation task, Price, Green, and Studnitz (1999) reported that German-English bilinguals showed increased activity in areas involved in attentional control (e.g., the anterior cingulate gyrus) among others. Recently, A. E. Hernandez (2009) found that participants' anterior cingulate gyrus and dorsolateral prefrontal cortex were highly active during language switching. These areas have also been associated with naming in a speaker' less proficient L2 (A. E. Hernandez & Meschyan, 2006). These studies offer a clue that parallel translation may be best explored using a secondary task that has been shown effective in occupying the CE of working memory in a dual-task paradigm. But an ideal secondary task should not involve verbal activity, as such a task could interfere with the primary reading task and also the process of parallel translation which could involve the language production system. One potentially useful task may be the tapping task (Gilhooly, Logie, & Wynn, 2002). The role that CE of WM might play during comprehension and production in sentence interpreting could be studied by occupying CE with the tapping task in selected stage(s) during sentence interpreting.

The finding that only participants' reading times in critical area, but not their memory performances in sentence translation and digit recall, were sensitive to the load and word order congruence can also be taken to support Gile's (2008a) approach of using smaller text unit, e.g., sentence, and local analysis to studying cognitive load-related limitations in language interpreting. Although Gile (2008a) originally applied this approach to simultaneous interpreting research, suggesting that interpreters' processing of a previous sentence could export processing load into the next sentence, the same approach could perhaps also be applied to study consecutive interpreting. Gile (2008a) proposed that even though his Effort Model emphasises that interpreters' total resource available is often exceeded by the total resource required, cognitive saturation does not necessarily occur at global level. Instead, processing deficit due to overload during language interpreting could be triggered by local processing difficulties. Take the results in Experiment 5 for example. Participants' memory performance in sentence translation did not differ across conditions, but their reading times were affected by word order congruence between SL and TL. To the extent that longer reading times indicate higher cognitive demand, Experiment 5 demonstrates Gile's (2008a) point that more resources injected into local processing were associated with longer reading times for local area of interest, but did not lead to performance difference at a global level, e.g., sentence memory. This was also true when participants were preloaded with digit list in Experiments 6 and 7. Because accessing the syntactic or lexical knowledge of a TL is a major component of language interpreting, the ease with which interpreters access their lexicons can, to a certain degree, determine their work load during grammatical encoding for TL production. Experiment 6 shows that when participants appeared to access TL syntactic information during comprehension, their speech rates in TL production were faster. But when their working memory was occupied by a secondary task, preventing them from accessing TL information during comprehension, they had to access TL information on the fly during TL production. In this case, it could be a pay-now or pay-later scenario if cognitive load in interpreters' access to TL information is seen as a debt. In other words, although early access to TL information can be resource-demanding (the debt is paid early on), it reduces the cognitive load in a subsequent process. Under conditions where early TL access is not possible, cognitive load deriving from conceptualisation, lexical access and grammatical encoding can exceed interpreters' capacity. In this pay-debt-later scenario, cognitive overload can have a manifestation of lower speech rate as shown in Experiment 6. There could also be other signs of disfluency, such as more filled pauses or more retracing. But further analyses for Experiments 5-7 are required to confirm these possibilities. In relation to methodological issues, Gile's (2008a) proposal and preliminary results of this thesis converge on using sentences as testing material and as the unit of analysis. In addition to common index of interpreting performance such as proportional proposition reproduced in a TL, the inclusion of online measures during decoding and recoding phases of interpreting will shed more light on interpreters' resource allocation in language interpreting.

5.2 Parallel Translation

By extending the paradigm of Ruiz et al. (2008), Experiment 5 provided supporting evidence for parallel translation in a sentence interpreting task, but the rest of experiments in the same series did not converge on this result. Two points of discussion in this section are 1) whether parallel translation was automatic; and 2) the implications of the findings in speech rate analysis in understanding the concept '*encoding*' in language interpreting.

If a significant interaction between word order congruency and recall method in Ruiz et al. (2008) and in Experiment 5 of this thesis can be taken as evidence for parallel translation, the lack of interaction between preload and word order congruency in the reading time data of Experiment 6 and 7 seemed to indicate that parallel translation was not employed by participants. Surprisingly, participants significantly reduced rather than increased their reading times when they were under heavy processing load. But more surprisingly, their interpreting performance as in the proportions of reproduced propositions was not affected by the experimental manipulation in digit preload or word order. They also performed equally well in digit recall across load conditions. Rather than automatic activation of TL properties during reading, these observations suggest that parallel translation may be strategic in nature: participants can flexibly allocate their attentional resources to ensure that concurrent tasks (reading for comprehension and memorising digits) were processed sufficiently and efficiently. Since there was no sign of decrement in terms of the quantity of the production data in load conditions relative to that in the no-load condition, parallel translation might not be an absolutely necessary stage in sentence interpreting. But these conclusions were based solely on the reading time data. Indeed, at least two questions can be asked: 1) what could motivate participants' parallel translation; and 2) whether reading time data can tease apart automatic and strategic processes of parallel translation?

First, given that participants were informed in advance of the task that their goal of reading was translation, there was a strong reason to suggest that this goal might drive their use of the strategy of parallel translation. Another motivation for using parallel translation could be that participants were translating into their L2, English. It has been assumed in second language production models that second language speakers have less complete lexical knowledge and have fewer automatised procedures in grammatical encoding (De Bot, 1992; Poulisse, 1997). It was possible that participants took the opportunity to retrieve lexical and syntactic information when they could. If this kind of active search in their TL knowledge can be seen as a form of advance planning, perhaps this effort could be reflected on timing-based measures such as speech onset latencies or speech rates.

A systematic analysis of speech rate revealed that participants' speech rates were significantly faster in translating incongruent than congruent sentences only in Experiment 6. Therefore, apart from the reading time data, there appeared to be additional evidence supporting the use of parallel translation during SL comprehension. But caution must be taken in interpreting this result. First of all, the analysis of speech rates did not show a consistent pattern across Experiments 5-7. Among the three experiments, only Experiment 6 showed a difference in speech rates in TL production. Moreover, this difference was significant only in by-subject but not in by-item analysis. Therefore this discussion is focused on Experiment 6.

If the speech rate difference were to be taken as complementary evidence for parallel translation in Experiment 6, it would require us to show that reading times were significantly different between two types of sentences, i.e. there should be an effect of word order congruency on reading times. However, reading times were significantly different between two types of sentences in no-load conditions while they were only numerically different in load conditions (5-digit and 7-digit preload). Thus the relationship between the two dependent variables was not straightforward. The argument so far has been based on the assumption that reading time difference implies parallel translation. But, it appears this assumption was a sufficient but not a necessary condition to implicate parallel translation. In other words, if it does not take significantly longer for an interpreter to access the TL information of an SL item than to simply comprehend it, then it would seem premature to conclude that parallel translation is strategic in nature. It is proposed that the speech rate could complement the reading time data in investigating the interface of interpreters' working languages during interpreting. Although the reading time data and the speech rate analysis did not converge on participants' use of parallel translation in

Experiment 6, the implications are discussed if the numerical differences of reading time are taken at face value. The following section first considers a few alternative accounts for faster speech rates in incongruent than congruent sentences before explaining as to why faster speech rate could be an indicator of early access to TL information.

In congruent conditions, sentences can be translated word-for-word between languages. If sentence interpreting can be seen as sentence recall in a different language, one would expect an SL sentence to prime its TL interpreting, whereby participants used the SL word order in constructing TL sentences (Potter & Lombardi, 1998). This priming effect of word order could lead to faster speech rates in congruent than incongruent conditions (e.g., MacKay & Bowman, 1969). However this was not substantiated. Instead, sentences that were processed longer were translated more fluently, suggesting that the structural persistence effect might have been overridden by the effect of parallel translation. The fact that relative clauses were produced significantly faster than were simple clauses, perhaps indicates that participants might have planned their translation during comprehension.

The higher speech rate in incongruent conditions cannot be explained by a frequency-based account either, since the number of trials were equal in congruent and incongruent conditions. Moreover, changing word order was compulsory for incongruent sentences, unlike those sentences typically used in the syntactic priming studies that have options, e.g. prepositional vs. double-object constructions. J. K. Bock and Loebell (1990) observed a trend of more fluent production for passive than active sentences. In their design, participants were free to alternate between two voices, so a frequency-based account needed to be considered. In the experiments in this thesis, neither congruent nor incongruent types of sentences received more repetitions in the experiment. Although it cannot be determined what this planning might involve, the suggestion that participants might have accessed the lexical and syntactic properties of a TL remains plausible (Ruiz et al., 2008). As mentioned earlier, advance planning could alleviate the process load on speakers' working memory, as early access to TL lexical knowledge could reduce the chances for word retrieval difficulty to occur during TL grammatical encoding. To be speculative, it is also possible participants could have linearised the accessed TL lexical items prior to sentence utterances.

The speech rate pattern in Experiment 6 might also imply a pay-now/pay-later trade-off between planning during SL comprehension and planning during TL production, particularly in retrieving the grammatical information in a TL. If the reading time difference in each preload condition is taken at face value, when participants were slower in reading incongruent than reading congruent sentences, their speech rates were faster for incongruent than congruent sentences. In other words, when they decided to plan on-line while speaking, it appeared to slow down their speech rate. To some degree, it implies a sequential (or vertical) translation when participants found it difficult to access TL information during comprehension, and recoding had to be delayed until TL production began. This sequential translation, however, can lead to slower speech rates of TL production. One caveat to this account was that it was not clear exactly at what stage the advance planning took place, because no record of the speech onset latencies in Experiment 6 was available. It was possible that participants were slower in starting to translate sentences for incongruent than for congruent sentences, i.e. they could have planned speech production during the interval between the end of sentence presentation and the start of TL production. Future research should address this issue.

Although it was predicted that a longer reading time would be associated with faster speech onset, no such difference was found in Experiment 7. This result seems to support incremental language production – speakers started delivering output as soon as a unit of speech plan was ready for articulation. Should this result generalise to Experiment 6, in such a scenario only speech rate, but not onset latencies, was sensitive to the digit preload, then the performance pattern in Experiment 6 would extend the incrementality of language production in spontaneous speech to sentence interpreting. According to F. Ferreira and Engelhardt (2006), incrementality of production can lead to efficient processing by spreading out the processing load over the entire utterance rather than having the load localised at the point of its initiation. If this argument extended to sentence interpreting in Experiment 6, it might account for the advantage of advance planning for incongruent sentence interpreting. Since parallel translation as a form of advance planning could alleviate processing load during grammatical encoding in TL production, the spread-out load across an utterance would become smaller for incongruent sentences than congruent sentences. One indicator for reduced load during language production could be a faster speech rate in Experiment 6. Again, this argument requires future research that takes reading times, onset latencies, and speech rates of sentence interpreting into consideration.

Taken together, it can be tentatively concluded that working memory was required in sentence interpreting during comprehension when parallel translation was employed, and the evidence from the speech rate data seemed to indicate that parallel translation as a form of advance planning could alleviate the process load on working memory during TL language production, leading to more fluent speech output. Although the relationship between the susceptibility of the advantage of parallel translation and external distractors (e.g., digit load during sentence production) cannot be decided yet, it could be an important finding if an interaction in speech rate between load and word order were found, i.e., speech rate only differed in the no-load, but not in the preload conditions. The implication of an interaction between load and word order in speech rate is that the product of interpreters' advance planning might require working memory to keep it active for later use. Unfortunately, the analysis showed that speech rates were not sensitive to word order or load, and they did not interact. Future research is required to determine whether digit load during TL production can remove the advantage of advance planning.

Given that the speech rate can be taken as a complementary measure in exploring parallel translation, it would be interesting to study the relationship between participants' bilingual competence and the strategy use of parallel translation. As mentioned in section 2.3, bilingual participants can be profiled in terms of their AoA and language learning context. When languages were *learnt* rather than *acquired*, many aspects of linguistic operations are thought to rely on the declarative memory, leading to qualitatively and quantitatively different processing (Paradis, 2009). The differences between proficient and less proficient bilinguals can be broken down by looking at four aspects of their mental lexicon as elaborated in section 2.3.3. Compared to proficient bilinguals, less proficient bilinguals might possess bilingual lexicons that show greater asymmetry in terms of lexicon size (Read, 2000), weaker connections of lexical items between and within each language, and less automatic lexical access (Meara, 1996), all of which could lead to different patterns of activation when a target language word is presented in a self-paced reading task. When the findings of language-nonselective access predict that a target-language word can discretely activate a proficient bilingual's non-target-language equivalent in a single word production task, the same may not be true of less proficient bilinguals. For less proficient bilinguals, access to the translation equivalent of a target word might "require additional ability to resolve cross-language competition from alternatives in L1 with similar meanings" (Kroll & Tokowicz, 2001). Because Ruiz et al. (2008) and Experiments 5-7 did not find evidence for co-activation of TL information in reading for later repetition, the question remains open as to whether language-non-selective access generalises to the processing of lexical access at the sentential level. Since the participants in Experiments 5-7 were not early bilinguals, nor were they professional interpreters, the results do not allow generalisation beyond these samples in this thesis. Future research should address factors of language proficiency and language dominance with measures of higher sensitivity.

There is evidence suggesting that highly proficient L2 speakers can be indistinguishable from native speakers in accent (Flege, MacKay, & Piske, 2002), and in implicit knowledge of L2 (Pallier et al., 2003), given a great deal of exposure to and practice of L2. Comparisons between monolingual and highly proficient late bilinguals, however, suggest that the two groups differ in parsing and production. Clahsen and Felser (2006) reported that proficient L2 speakers relied less on syntactic information than L1 monolinguals. Sorace (2003) compared English-Italian bilinguals with Italian monolinguals and showed bilinguals demonstrated optionality in pronoun drop in the context where monolinguals clearly preferred a construction with a null pronoun. It appears that the difference between monolingual and bilingual processing lies in the integration of syntactic knowledge and knowledge from other domains (Jakubowicz & Nash, 2001; Sorace, 2005). Sorace (2005) contends that constructions are more complex when they require integration of knowledge of syntax and other domains, e.g., pragmatic, semantic. When bilinguals' two languages differ in their syntactic complexity, optionality occurs when one language instantiates more economical options. But it remains to be tested whether optionality in sentence construction is intrinsic to bilinguals' interpretation that is constrained by their linguistic competence or it is constrained by their limited resource for the computation involved in the integration of knowledge across domains and levels. The selective optionality among highly proficient English-Italian bilinguals (Sorace, 2003) and morphological problems such as inflection and case marking (Lardiere, 1998; Prvost & White, 2000) in L2 could as well be derived from computation problems. Even when professional interpreters' language proficiency is at the highest spectrum, they were reported to show subject-verb agreement errors and loss of reference in language interpreting (Agrifoglio, 2004). If the computation account holds, future studies should consider higher-order processes in exploring issues revolving the representation and processing in language interpreting. Given that interpreters are supplied with a strong task and contextual cue towards their bilingual mode, experiments using highly proficient bilinguals and sensitive online measures (see the next section for more details regarding methodologies) can test the generalisability of the results in Experiments 5-7 to bilinguals whose language proficiency is at the highest spectrum.

As Ruiz et al. (2008) and Experiments 5-8 in this thesis required participants to translate only from their L1 into L2 only, more research is needed to further test the translation hypotheses by replicating the same procedure in the reverse direction of translation, e.g., English into Chinese.

5.3 Methodological Considerations

Behavioural data in Ruiz et al. (2008) and the experiments in this thesis have provided preliminary evidence supporting the psychological reality of parallel translation. Although the self-paced reading paradigm showed its usefulness, one major criticism is that reading times do not reveal the source of the changes in processing because the reading time of a certain word can have several components – "they could be additive and serial, they might interact, or they might operate in parallel" (Haberlandt, 1994, p. 9). Take Experiment 6 for example, if the faster speech rate in TL production were a consequence of parallel translation during SL comprehension, then reading times should show differences between congruent and incongruent sentences. However this was not the case. Reading times only differed significantly in no-load condition, whereas they were only numerically different in load conditions. Of course it would be premature to arrive at a conclusion about the suitability of reading time before more research is carried out in exploring different language pairs, translation direction, and sentence structures.

Another source of criticism might stem from the fact that the self-paced reading task allows participants to pace the input rate of to-be-translated material, contrasting the high speech rate in spoken language in day-to-day life. While the self-paced reading paradigm can be used to further examine translation hypotheses, at the same time other methodologies could be used to address the same research questions and see if different methodologies converge on the same results. Ibáñez et al.'s (2008) ERPs study was a promising start in studying the process of interpreting. But even when a neurophysiological paradigm is not immediately accessible, researchers could explore whether a change of input modality can replicate the findings of Macizo and Bajo (2006) and Ruiz et al. (2008). In this regard, the auditory moving window technique can be considered (F. Ferreira, Henderson, Anes, Weeks Jr., & McFarlane, 1996). In the auditory moving window paradigm, sentences are naturally spoken, recorded and digitised. Participants press a button to listen to a sentence word-for-word just like they do in self-paced reading. The dependent variable is the duration required for one word or one phrase. A major difference between visual and auditory presentation is that words in self-paced reading paradigm stay on the screen before participants press the button to reveal the next word/phrase, whereas words in auditory moving window paradigm can be played only once for a very short period of time. It is not clear how this difference in sentence presentation would affect the way participants process sentences for translation. But if it was true that words which are permanently present and available make translation an 'open prey' for transcoding (De Groot, 1997), the transient presentation of words/phrases in the auditory moving window paradigm would probably show different results from those in self-paced reading studies. Regarding the technique, F. Ferreira et al. (1996) and Waters and Caplan (2002) have provided results that support the validity and usefulness of this task. One complication intrinsic to the auditory moving window technique that could make material preparation a little more difficult than self-paced reading is the control and manipulation of the prosody of sentences. Nevertheless, this technique seems to be a reasonably promising step for researchers to take, in exploring the ecological validity of parallel translation.

Indeed, Macizo and Bajo (2006) and Ruiz et al. (2008) have opened up an new avenue to investigate the way in which bilinguals use their knowledge of languages in a complex task such as interpreting. The idea of accessing the TL properties during SL comprehension in language interpreting is plausible but its relevance to and significance in theories of bilingualism requires rigorous investigation. For example, intuitively it would be hard to imagine that interpreters access the TL information for every single item in processing a to-be-translated text. An important question awaiting empirical test is how bilinguals decide when and what type of information they access the TL information when processing an SL text for later interpreting. This question has received partial treatment in Ruiz et al. (2008). They established that the type of information that participants accessed was associated with lexical frequency, but it appeared that participants did not access the TL information for each item in an SL text. If one was to further explore the type of TL information that becomes activated, perhaps one could use the cross-modal lexical priming task (CMLP) (e.g., A. E. Hernandez, Bates, & Avila, 1996; Heredia & Stewart, 2002) to investigate whether bilinguals retrieve the TL phonological information for an SL word and whether the activated information is maintained by a mechanism such as rehearsal until TL production is required. Typically, participants listen to a sentence while watching a fixation point on a computer screen. At some point (usually immediately after a prime word), the sentence presentation is interrupted and the fixation point is replaced by a target word. At this moment, participants are required to make a timed response, e.g., lexical decision or naming. Their response times to experimental and control items will be compared to decide whether priming took place. A. E. Hernandez et al. (1996) showed that priming effect was modulated by semantic congruency (semantically related vs. unrelated), linguistic congruency (within- or cross-language) and predictability (blocked vs. mixed design). In addressing the question of whether participants access the TL phonological information for an SL item in this CMLP paradigm, one could vary the phonological similarity of target words to the translation equivalent while keeping the linguistic and semantic congruency constant. If indeed participants access the phonological information of the translation equivalent for an SL item, one would expect to find a facilitatory effect on naming latencies when the target words are phonologically similar to the targets (e.g., Vitevitch, 2002). In addition to the response times, participants' production data could also be useful in elucidating whether early access to TL information (if it occurs) has any influence on the incrementality of TL production, measured by the onset latencies and indicators typically used to index speech fluency and speech rate.

5.4 Conclusions

A full consideration of the components of translation has guided this thesis in formulating its research questions of whether working memory was involved in language interpreting, and in what way, if it was. The first series of experiments explored the hypothesis that working memory was shared by memory maintenance and grammatical encoding during language production in discourse interpreting. Although the material confounded the results, Experiment 1-4 highlighted the importance of comprehension in interpreting, and complexity in selecting long passages for experimental testing. The same research questions were explored with another series of experiments with modifications of the hypotheses and design. Working memory was assumed as a workspace that interfaced bilinguals' two languages. By repeating the procedure of Ruiz et al. (2008), Experiment 5 replicated their findings which seemed to imply the use of parallel translation. In Experiment 6 and 7, this paradigm was combined with the dual-task technique (Waters et al., 2003). Although the performance patterns were unexpected, evidence suggested that working memory was required when participants appeared to use parallel translation when resources were available to them. But when their working memory was occupied by a secondary task, resources were flexibly allocated to concurrent tasks, resulting in no trade-off in performance. This result supported the view of working memory as task-specific as opposed to task-general resource pool. An analysis of speech rate provided additional evidence that suggested participants' advantage in speech rate during TL production might be a result of parallel translation during SL comprehension. Parallel translation might be a strategy used to reduce process load in the production stage. The role of working memory in how the advantage of parallel translation can be fully demonstrated by participants is a theoretically relevant question but remains to be examined.

References

Acheson, D. J., & MacDonald, M. C. (2009). Verbal working memory and language production: Common approaches to the serial ordering of verbal information. *Psychological Bulletin*, *135*, 50–68.

Agrifoglio, M. (2004). Sight translation and interpreting: A comparative analysis of constrints and failures. *Interpreting*, *6*(1), 43-67.

Alexieva, B. (2002). A typology of interpreter-mediated events. In F. Pochhacker & M. Shlesinger (Eds.), *The interpreting studies reader* (pp. 218–233). Routledge.

Allport, A., Styles, E. A., & Hsieh, S. (1994). Shifting attentional set: Exploring the dynamic control of tasks. In C. Umilta & M. Moscovitch (Eds.), *Attention and performance XV* (pp. 421–452). Cambridge, MA: MIT Press.

Amos, D. (2009, November). *Reporter's notebook: Taliban translator was a New Yorker.* Retrieved from `http://abcnews.go.com/WNT/story?id=129932&page=1`

Anderson, L. (1994). Simultaneous interpretation: Contextual and translation aspects. In *Bridging the gap: Empirical research in simultaneous interpretation.* (pp. 101–120). Amsterdam: John Benjamins.

Ardila, A. (2003). Language representation and working memory with bilinguals. *Journal of Communication Disorders*, *36*, 233–240.

Baddeley, A. D. (1978). The trouble with levels: A reexamination of Craik and Lockhart's framework for memory research. *Psychological Review*, *16*, 139–152.

Baddeley, A. D. (1983). Working memory. *Philosophical Transactions of the Royal Society of London, Series B*, *302*, 311–324.

Baddeley, A. D. (1986). *Working memory.* Oxford: Oxford University Press.

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

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

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

Baddeley, A. D. (2003). Working memory and language: An overview. *Journal of Communication Disorders*, *36*, 189–208.

Baddeley, A. D. (2007). *Working memory, thought, and action.* Oxford: Oxford University Press.

Baddeley, A. D., Bressi, S., Della Sala, S., Logie, R. H., & Spinnler, H. (1991). The decline of working memory in Alzheimer's Disease: A longitudinal study. *Brain*, *114*, 2521–2542.

Baddeley, A. D., Chincotta, D., & Adlam, A. (2001). Working memory and the control of action: evidence from task switching. *Journal of Experimental Psychology: General*, *130*, 641–657.

Baddeley, A. D., Eldridge, M., & Lewis, V. (1981). The role of subvocalisation in reading. *Quarterly Journal of Experimental Psychology*, *33A*, 439–454.

Baddeley, A. D., & Hitch, G. (1974b). Working memory. *Psychology of learning and motivation*, *8*, 47–89.

Baddeley, A. D., & Hitch, G. J. (1974a). Working memory. In G. Bower (Ed.), *Recent advances in learning and motivation.* (Vol. III, pp. 47–90). New York:

Academic Press.

Baddeley, A. D., Hitch, G. J., & Allen, R. J. (2009). Working memory and binding in sentence recall. *Journal of Memory and Language*, *61*, 438–456.

Baddeley, A. D., Lewis, V., & Vallar, G. (1984). Exploring the articulatory loop. *The Quarterly Journal of Experimental Psychology*, *36A*, 233-252.

Baddeley, A. D., Lewis, V. J., Eldridge, M., & Thomson, N. (1984). Attention and retrieval from long-term memory. *Journal of Experimental Psychology: General*, *113*, 518–540.

Baddeley, A. D., Logie, R. H., Nimmo-Smith, I., & Brereton, N. (1985). Components of fluent reading. *Journal of Memory and Language*, *24*, 119–131.

Baddeley, A. D., Papagno, C., & Vallar, G. (1988). When long-term learning depends on short-term storage. *Journal of Memory and Language*, *27*, 586-595.

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.

Baddeley, A. D., & Wilson, B. (1985). Phonological encoding and short-term memory in patients without speech. *Journal of Memory and Language*, *24*, 490–502.

Baetens-Beardsmore, H., & Lee, E. J. (1975). *On bilingual competence* (Tech. Rep.). Rapport d'Activites de l'Institut de Phonetique.

Bajo, M. T., Padilla, F., & Padilla, P. (2000). Comprehension processes in simultaneous interpreting. In C. A., N. G. San Salvador, & Y. Gambier (Eds.), *Translation in context* (pp. 127–142). John Benjamins.

Bajo, M. T., Padilla, P., Munoz, R., Padilla, F., Gomez, C., Puerta, M. C., et al. (2001). Comprehension and memory processes in translation and interpreting. *Quaderns: Revista de traducció*, *6*, 27–31.

Barik, H. C. (1972). Interpreters talk a lot, among other things. *Babel*, *18*, 3–10.

Barik, H. C. (1975). Simultaneous interpretation - Qualitative and linguistic data. *Language and Speech*, *18*, 272–297. (Bg253 Times Cited:9 Cited References Count:13)

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General*, *133*, 83–100.

Basnight-Brown, D. M., & Altarriba, J. (2007). Differences in semantic and translation priming across languages: The role of language direction and language dominance. *Memory and Cognition*, *35*, 953–965.

Basnight-Brown, D. M., Chen, L., Hua, S., Kostić, A., & Feldman, L. B. (2007). Monolingual and bilingual recognition of regular and irregular english verbs: Sensitivity to form similarity varies with first language experience. *Journal of Memory and Language*, *57*, 65–80.

Bell, R. T. (1991). *Translation and translating: Theory and Practice*. London: Longman.

Bernolet, S., Hartsuiker, R. J., & Pickering, M. J. (2007). Shared syntactic representations in bilinguals: Evidence for the role of word-order repetition. *Journal Of Experimental Psychology: Learning, Memory and Cognition*, *33*, 931–949.

Bever, T. G. (1970). The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the development of language.* (pp. 279–362). New York: Wiley.

Bierwisch, M. (1969). On certain problems of semantic representation. *Foundations of Language*, *5*, 153–184.

Bierwisch, M., & Schreuder, R. (1992). From concepts to lexical items. *Cognition*, *42*, 23–60.

Birdsong, D. (1992). Ultimate attainment in second language acquisition. *Language*, *68*, 706–755.

Birdsong, D. (2006). Age and second language acquisition and processing: A selective overview. *Language Learning*, *56*, 9–49.

Birdsong, D., & Molis, M. (2001). On the evidence for maturational constraints in second-language acquisition. *Journal of Memory and Language*, *44*, 235–249.

Birdsong, D., & Paik, J. (2008). Second language acquisition and ultimate attainment. In B. Spolsky & F. Hult (Eds.), *The handbook of educational linguistics.* (pp. 424–436). Blackwell.

Black, J. (1981). The effects of reading purpose on memory. In J. Long & A. Baddeley (Eds.), *Attention and performance IX: Proceedings of the ninth international symposium on attention and performance, Jesus College, Cambridge, England, July 13-18, 1980* (pp. 347–362). Philadelphia: Lawrence Erlbaum Associates.

Bock, J. K. (1977). The effect of pragmatic presupposition on syntactic structure in question answering. *Journal of Verbal Learning and Verbal Behavior*, *16*, 723–734.

Bock, J. K. (1996). Language production: Methods and methodologies. *Psychonomic Bulletin and Review*, *3*, 395–421.

Bock, J. K., & Irwin, D. E. (1980). Syntactic effects of information availability in sentence production. *Journal of Verbal Learning and Verbal Behavior*, *19*, 467–484.

Bock, J. K., & Loebell, H. (1990). Framing sentences. *Cognition*, *35*, 1–39.

Bock, J. K., & Warren, R. K. (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition*, *21*, 47–67.

Bock, K., & Cutting, J. C. (1992). Regulating mental energy: Performance units in language production. *Journal of memory and language*, *31*, 99–127.

Boersma, P., & Weenink, D. (2001). Praat: A system for doing phonetics by computer. *Glot International.*, *5*, 341–345.

Booth, A. D. (1958). *Aspects of translation.* London: Secker and Warburg.

Bourke, P. A., Duncan, J., & Nimmo-Smith, I. (1996). A general factor involved in dual-task performance decrement. *Quarterly Journal of Experimental Psychology*, *49A*, 525–545.

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. *Cognition*, *75*(2), 13–25. (311HX Times Cited:91 Cited References Count:30)

Bransford, J., & Jeffery, J. (1971). The abstraction of linguistic ideas. *Cognitive Psychology*, *2*, 331-350.

Bransford, J. D., Barclay, J. R., & Franks, J. J. (1972). Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology*, *3*, 193–209.

Broeck, R. Van den. (1991). The general model of translation process. In K. M. van L-Z & T. Naaijkens (Eds.), *Translation studies: The state of the art – Proceedings of the First James S Holmes Symposium on Translation Studies* (pp. 105–114). Amsterdam: Rodopi.

Broek, P. V. den, Lorch, R. F., Linderholm, T., & Gustafson, M. (2001). The effects of readers' goals on inference generation and memory for texts. *Memory & Cognition*, *29*, 1081–1087.

Butterworth, B. (1975). Hesitation and semantic planning in speech. *Journal of Psycholinguistic Research*, *4*, 75–87.

Campbell, S. (1998). *Translation into the second language.* London: Longman.

Campbell, S., & Wakim, B. (2007). Methodological questions about translation research: A model to underpin research into the mental processes of translation. *Target*, *19*, 1–19.

Caplan, D., & Waters, G. S. (1995). On the nature of the phonological output planning processes involved in verbal rehearsal: Evidence from aphasia. *Brain and Language*, *48*, 191–220.

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

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

Carroll, J. B., & White, M. N. (1973). Word frequency and age-of-acquisition as determiners of picture-naming latency. *Quarterly Journal of Experimental Psychology*, *25*, 85–95.

Chang, C., & Schallert, D. (2007). The impact of directionality on Chinese-English simultaneous interpreting. *Interpreting*, *9*, 137–176.

Chen, H. C. (1992). Reading comprehension in Chinese: Implications from character reading times. In H.-C. Chen & O. J. L. Tzeng (Eds.), *Language processing in chinese.* (pp. 175–205). Amsterdam: North-Holland.

Chen, H.-H. (1992). The transfer of anaphors in translation. *Literary and Linguistic Computing*, *7*, 231–238.

Cheung, A. (2007). The cantonese dative construction: Implications for processing. In *Programs of Conference on Ditransitive Constructions in Max Planck Institue for Evolutionary Anthropology, Leipzig. 23-25 November.*

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, Mass: M.I.T. Press.

Christoffels, I. K. (2006). Listening while talking: The retention of prose under articulatory suppression in relation to simultaneous interpreting. *European Journal of Cognitive Psychology*, *18*, 206–220.

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

Christoffels, I. K., & De Groot, A. M. B. (2005). Simultaneous interpreting: A cognitive perspective. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 454–479). New York: Oxford University Press.

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

Christoffels, I. K., Firk, C., & Schiller, N. O. (2007). Bilingual language control: An event-related brain potential study. *Brain Research*, *1147*, 192–208.

Clahsen, H., & Felser, C. (1996). Grammatical processing in first and second language learners. *Applied Psycholinguistics*, *27*, 3–42.

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, *27*, 3-42.

Clark, H. E., & Clark, E. V. (1977). *Psychology and language: An introduction to psycholinguistics.* San Diego: Harcourt Brace Jovanovich.

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

Cocchini, G., Logie, R., Della Sala, S., MacPherson, S., & Baddeley, A. D. (2002). Concurrent performance of two memory tasks: Evidence for domain-specific working memory systems. *Memory & Cognition*, *30*, 1086–1095.

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

Comrie, B. (1989). *Language universals and linguistic typology: Syntax and morphology.* Chicago: The University of Chicago Press.

Connor, U. (1984). Recall of text: Differences between first and second language readers. *TESOL Quarterly*, *18*, 239–256.

Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology*, *55*, 429-432.

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 user's guide. *Psychonic Bulletin & Review*, *12*, 769–786.

Cook, V. J. (1975). Strategies in the comprehension of relative clauses. *Language and Speech*, *18*, 204–212.

Coppieters, R. (1987). Competence differences between native and near-native speakers. *Language*, *63*, 544–573.

Corley, M., & Scheepers, C. (2002). Syntactic priming in english sentence production: categorical and latency evidence from an internet-based study. *Psychonomic Bulletin & Review*, *9*, 126–131.

Costa, A. (2004). Speech production in bilinguals. In T. K. Bhatia & W. C. Ritchie (Eds.), *Handbook of Bilingualism* (pp. 201–223). Oxford: Oxford University Press.

Costa, A., & Caramazza, A. (1999). Is lexical selection in bilingual speech production language-specific? *Bilingualism: Language and Cognition*, *2*, 231–244.

Costa, A., Miozzo, M., & Caramazza, A. (1999). Lexical selection in bilinguals: Do words in the bilingual's two lexicons compete for selection? *Journal of Memory and Language*, *41*, 365–397.

Council, B. (2009, Novermber). *What is IELTS?* Retrieved from `http://www.britishcouncil.org/india-exams-ielts-about.htm`

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior.*, *11*, 671–684.

Crookes, G. (1989). Planning and interlanguage variation. *Studies in Second Language Acquisition*, *11*, 367–383.

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

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

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

Daneman, M., & Hannon, B. (2007). What do working memory span tasks like reading span really measure? In N. Osaka, R. H. Logie, & M. D'Esposito (Eds.), *The cognitive neuroscience of working memory.* (pp. 21–42). Oxford, UK: Oxford University Press.

Danks, J. H., & Griffin, J. (1997). Reading and translation – a psycholinguistic perspective. In J. H. Danks, G. M. Shreve, S. B. Fountain, & M. K. McBeath (Eds.), *Cognitive processes in translation and interpreting.* (pp. 161–175). SAGE Publications.

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

Daro, V., Lambert, S., & Fabbro, F. (1996). Concscious monitoring of attention during simultaneous interpretation. *Interpreting*, *1*, 101–124.

De Bot, K. (1992). A bilingual production-model: Levelt's 'speaking' model adapted. *Applied Linguistics*, *13*, 1–24.

De Bot, K. (2000). Simultaneous interpreting as language production. In B. Englund & K. Hyltenstam (Eds.), *Language processing and simultaneous interpreting* (pp. 65–88). Amsterdam: Benjamins.

De Bot, K. (2003). Bilingual speech: From concepts to articulation. In K. Aguado (Ed.), *Themenschwerpunkt: Mündliche produktion in der fremdsprache* (pp. 92–103). Tübingen: Narr Verlag.

De Bot, K., & Schreuder, R. (1993). Word production and the bilingual lexicon. In R. Schreuder & B. Weltens (Eds.), *The bilingual lexicon.* (pp. 191–214). Amsterdam: John Benjamins.

De Groot, A. M. B. (1984). Primed lexical decision: Combined effects of the proportion of related prime-target pairs and the stimulus onset asynchrony of prime and target. *Quarterly Journal of Experimental Psychology*, *36A*, 253–280.

De Groot, A. M. B. (1997). The cognitive study of translation and interpretation: three approaches. In J. Danks, G. Shreve, S. Fountain, & M. McBeath (Eds.), *Cognitive processes in translation and interpreting* (pp. 25–76). Thousand Oaks: SAGE Publications.

De Groot, A. M. B., & Nas, G. L. J. (1991). Lexical representation of cognates and noncognates in compound bilinguals. *Journal of Memory and Language*, *30*, 90–123.

De Groot, A. M. B., & Poot, R. (1997). Word translation at three levels fo proficiency in a second language: The ubiquitous involvement of conceptual memory. *Langauge Learning*, *47*, 215-264.

De Jong, N. H., & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods*, *41*, 385–390.

DeKeyser, R. M. (2001). Automaticity and automatization. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 125–151). Cambridge: Cambridge University Press.

Den Heyer, K., Briand, K., & Dannenberg, G. (1983). Strategic factors in a lexical-decision task: evidence for automatic and attention-driven processes. *Memory & Cognition*, *11*, 374–381.

Di Giulio, D. V., Seidenberg, M., O'Leary, D. S., & Raz, N. (1994). Procedural and declarative memory: a developmental study. *Brain and Cognition*, *25*, 79–91.

Dijkstra, A., & Van Heuven, W. J. B. (1998). The bia model and bilingual word recognition. In J. Grainger & A. M. Jacobs (Eds.), *Localist connectionist approaches to human cognition* (pp. 189–225). Lawrence Erlbaum.

Dijkstra, T., Grainger, J., & Van Heuven, W. J. B. (1999). Recognition of cognates and interlingual homographs: The neglected role of phonology. *Journal of*

Memory and Language., *41*, 496–518.

Dijkstra, T., & Van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, *5*, 175–197.

Diller, H.-J., & Kornelius, J. (1978). *Linguistic probleme der Übersetzung.* Tubingen: Niemeyer.

Dillinger, M. (1994). Comprehension during interpreting: What do interpreters know that biinguals don't? In S. L. Moser-Mercer & Barbara (Eds.), *Bridging the gap: Empirical research in simulaneous interpretation.* (pp. 155–189). Amsterdam: John Benjamins.

Dixon, R., Hultsch, D., Simon, E., & Von Eye, A. (1984). Verbal ability and text structure effects on adult age differences in text recall. *Journal of Verbal Learning and Verbal Behavior*, *23*, 569–578.

Donovan, C. (2004). European masters project group: Teaching simultaneous interpretation into a b language: Preliminary findings. *Interpreting*, *6*, 205-216.

Dornyei, Z., & Kormos, J. (1998). Problem-solving mechanisms in L2 communication: A psycholinguistic perspective. *Studies in Second Language Acquisition*, *20*, 349–385.

Driessen, C. (1993). *Het tweetalige lexicon.* University of Nijmegen.

Duff, S., & Logie, R. H. (1999). Storage and processing in visuo-spatial working memory. *Scandinavian Journal of Psychology*, *40*, 251–259.

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

Duyck, W., & Brysbaert, M. (2004). Forward and backward number translation requres conceptual mediation in both balanced and unbalanced bilinguals. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 889–906.

Duyck, W., Vanderelst, D., Desmet, T., & Hartsuiker, R. J. (2008). The frequency effect in second-language visual word recognition. *Psychonomic Bulletin & Review*, *15*, 850–855.

Ecke, P. (2009). The tip-of-the-tongue phoenomenon as a window on (bilingual) lexical retrieval. In A. Pavlenko (Ed.), *The bilingual mental lexicon* (pp. 185–208). Multilingual Matters.

Engle, R. (2009, November). *Attention and working memory lab: Tasks.* Retrieved from `http://psychology.gatech.edu/renglelab/Tasks.htm`

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

Fathman, A. K. (1980). Repetition and correction as an indication of planning and execution processes among second language learners. In H. W. Dechert & M. Raupach (Eds.), *Towards a crosslinguistic assessment of speech production.* (pp. 77–86). Frankfurt: Peter Lang.

Favreau, M., & Segalowitz, N. S. (1983). Automatic and controlled processes in the first- and second -language reading of fluent bilinguals. *Memory & Cognition*, *11*, 565–574.

Fayol, M., Largy, P., & Lemaire, P. (1994). Cognitive overload and orthographic errors: When cognitive overload enhances subject–verb agreement errors. a study in french written language. *The Quarterly Journal of Experimental Psychology Section A*, *47*, 437–464.

Fedorenko, E., Gibson, E., & Rohde, D. (2006). The nature of working memory capacity in sentence comprehension: Evidence against domain-specific working memory resources. *Journal of Memory and Language*, *54*, 541–553.

Fedorenko, E., Gibson, E., & Rohde, D. (2007). The nature of working memory in linguistic, arithmetic and spatial integration processes. *Journal of Memory and Language*, *56*, 246–269.

Ferré, P., Sanchez-Casas, R., & Guasch, M. (2006). Can a horse be a donkey? Semantic and form interference effects in translation recognition in early and late proficient and nonproficient spanish-catalan bilinguals. *Language Learning*, *56*, 571–608.

Ferreira, F. (1991). Effects of length and syntactic complexity on initiation times for prepared utterances. *Journal of Memory and Language*, *30*, 210–233.

Ferreira, F., & Engelhardt, P. E. (2006). Syntax and production. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of psycholinguistics* (pp. 61–91). Elsevier.

Ferreira, F., Henderson, J., Anes, M., Weeks Jr., P., & McFarlane, D. (1996). Effects of lexical frequency and syntactic complexity in spoken-language comprehension: evidence from the auditory moving-window technique. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*, *22*, 324–335.

Ferreira, F., & Henderson, J. M. (1998). Linearization strategies during language production. *Memory & Cognition*, *26*, 88–96.

Ferreira, F., & Swets, B. (2002). How incremental is language production? Evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language*, *46*, 57–84.

Ferreira, V. S., & Pashler, H. (2002). Central bottleneck influences on the processing stages of word production. *Journal Of Experimental Psychology: Learning, Mememory and Cognition*, *28*, 1187–1199.

Fillmore, C. J. (1968). The case for case. In E. Black & R. T. Harms (Eds.), *Universals of linguistic theory.* (pp. 1–88). Holt: Reinhart, & Winston.

Finkbeiner, M., Almeida, J., Janssen, N., & Caramazza, A. (2006). Lexical selection in bilingual speech production does not involve language suppression. *Journal of Experimental Psychology: Learning Memory and Cognition*, *32*, 1075-1089.

Finkbeiner, M., Gollan, T. H., & Caramazza, A. (2006). Lexical access in bilingual speakers: What's the (hard) problem? *Bilingualism: Language and*

Cognition, *9*, 153–166.

Flege, J. E., MacKay, I. R. A., & Piske, T. (2002). Assessing bilingual dominance. *Applied Psycholinguistics*, *23*, 567–598.

Foote, R. (2010). Age of acquisition and proficiency as factors in language production: Agrement in bilinguals. *Bilingualism: Language and Cognition*, *13*, 99–118.

Fox, E. (1996). Cross-language priming from ignored words: evidence for a common representational system in bilinguals. *Journal of Memory and Language*, *35*, 357–370.

Frazier, L. (1978). *On comprehending sentences: Syntactic parsing strategies.* Unpublished doctoral dissertation, Storrs: University of Connecticut.

Frazier, L. (1995). Constraint satisfaction as a theory of sentence processing. *Journal of Psycholinguistic Research*, *24*, 437–468.

Fredriksson, A. (2000). Maze learning and motor activity deficits in adult mice induced by iron exporsure during a critical postnatal period. *Developmental Brain Research*, *119*, 65–74.

Frenck, C., & Pynte, J. (1987). Semantic representation and surface forms: A look at across-language priming in bilinguals. *Journal of Psycholinguistic Research*, *16*, 383–396.

Friedman, N., & Miyake, A. (2004). The reading span test and its predictive power for reading comprehension ability. *Journal of Memory and Language*, *51*, 136–158.

Friesen, D. C., & Jared, D. (2007). Cross-language message- and word-level transfer effects in bilingual text processing. *Memory & Cognition*, *35*, 1542–1556.

Frost, R. (1998). Toward a strong phonological theory of visual word recognition: true issues and false trails. *Psychologial Bulletin*, *123*, 71-99.

Garrett, M. F. (1982). Production of speech: Observations from normal and pathological language use. In A. W. Ellis (Ed.), *Normality and pathology in* *cognitive functions.* (pp. 19–76). New York: Academic Press.

Gathercole, S. E., & Baddeley, A. D. (1993). *Working memory and language.* Hove: Erlbaum.

Gathercole, S. E., Frankish, C. R., Pickering, S. J., & Peaker, S. (1999). Phonotactic influences on short-term memory. *Journal of Experimental Psychology: Learning Memory and Cognition*, *25*, 84–95.

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

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*, 1–76.

Gideon, T., Anthony, P., Miriam, S., & Daniel, S. (2008). *Beyond descriptive translation studies: investigations in homage to Gideon Toury.* Amsterdam: John Benjamins.

Gile, D. (1990). Scientific research vs. personal theories in the investigation of interpretation. In L. Gran & C. Taylor (Eds.), *Aspects of applied and experimental research on conference interpretation.* (pp. 28–41). Udine: Campanotto.

Gile, D. (1991). The processing capacity issue in conference interpretation. *Babel*, *37*, 15–27.

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

Gile, D. (2004). Translation research versus interpreting research: Kinship, differences and prospects for partnership. In C. Schèaffner (Ed.), *Translation research and interpreting research : traditions, gaps and synergies* (pp. 10–34). Multilingual Matters.

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

Gile, D. (2008b, December). *Theories and models for research, theories, and models for training.* Sorbonne Nouvelle. Retrieved from `http://montereyjournal.net/home/wp-content/uploads/2008/12/microsoft-word-gile.pdf`

Gilhooly, K. J., & Logie, R. H. (1982). Word age-of-acquisition and lexical decision making. *Acta Psychologica*, *50*, 21–34.

Gilhooly, K. J., Logie, R. H., & Wynn, V. E. (2002). Syllogistic reasoning tasks and working memory: Evidence from sequential presentation of premises. *Current Psychology*, *21*, 111–120.

Glenberg, A. M., Smith, S. M., & Green, C. (1977). Type 1 rehearsal: Maintenance and more. *Journal of Verbal Learning and Verbal Behavior*, *16*, 339–352.

Glisan, E. W. (1985). The effect of word order on listening comprehension and pattern retention: An experiment in spanish as a foreign language. *Language Learning*, *35*, 443–472.

Gollan, T. H., & Acenas, L.-A. R. (2004). What is a tot? Cognate and translation effects on tip-of-the-tongue states in spanish-english and tagalog-english bilinguals. *Journal of Experimental Psychology: Learning, Memory and Cognition.*, *30*, 246–269.

Gollan, T. H., Montoya, R. I., Fennema-Notestine, C., & Morris, S. K. (2005). Bilingualism affects picture naming but not picture classification. *Memory & Cognition*, *33*, 1220–1234.

Grabner, R. H., Brunner, C., Leeb, R., Neuper, C., & Pfurtscheller, G. (2007). Event-related EEG theta and alpha band oscillatory responses during language translation. *Brain Research Bulletin*, *72*, 57–65.

Grainger, J., & Beauvillain, C. (1988). Associative priming in bilinguals: Some limitations of interlingual facilitation effects. *Canadian Journal of Psychology*, *42*, 261–273.

Green, D. (1993). Towards a model of L2 comprehension and production. In *The bilingual lexicon.* (pp. 249–278). Amsterdam: John Benjamins.

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

Gries, S., & Wulff, S. (2005). Do foreign language learners also have constructions? Evidence from priming, sorting and corpora. *Annual Review of Cognitive Linguistics*, *3*, 182–200.

Grosjean, F. (1997). Processing mixed language: Issues, findings, and models. In A. M. B. de Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 225–254). Lawrence Erlbaum.

Grosjean, F. (2001). The bilingual's language modes. In J. Nicol (Ed.), *One mind, two languages: Bilingual language processing.* (pp. 1–22). Blackwell.

Grosjean, F. (2008). *Studying bilinguals.* Oxford: Oxford University Press.

Gutt, E.-A. (1991). *Translation and relevance: Cognition and context.* Oxford: Blackwell.

Haberlandt, K. (1994). Methods in reading research. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics.* (pp. 1–31). Academic Press.

Hakuta, K., Gould, L., Malakoff, M., Rivera, M., & Rodriguez-Landsberg, M. (1988). Translation and interpretation in bilingual puerto-rican students. Paper presented at AERA. New Orleans.

Hakuta, K., & Malakoff, M. (1987). Translation skills in bilingual children. Paper presetned at the society for research in child development. Baltimore.

Harris, B. (1977). The importance of natural translation. In B. Harris (Ed.), *Working papers on bilingualism 12* (pp. 96–114). School of Translators and Interpreters, The University of Ottawa.

Harris, B. (1980). How a three-year-old translates. In *Patterns of bilingualism. RELC anthology series 8.* Singapore: National University of Singapore Press.

Harris, B., & Sherwood, B. (1978). Translating as an innate skill. In D. Gerver & H. Sinaiko (Eds.), *Langauge, Interpretation and Communication.* (pp. 155–170). Plenum Press.

Harris, D. (1970). Report on an experimental group-administered memory span test. *TESOL Quarterly*, *4*, 203–213.

Hartley, J. T. (1993). Aging and prose memory: Tests of the resource-deficit hypothesis. *Psychology and Aging*, *8*, 538–551.

Hartsuiker, R. J., & Barkhuysen, P. N. (2006). Language production and working memory: The case of subject-verb agreement. *Language and Cognitive Processes*, *21*, 181–204.

Hartsuiker, R. J., Kolk, H. H. J., & Huinck, W. J. (1999). Agrammatic production of subject-verb agreement: The effect of conceptual number. *Brain and language*, *69*, 119–160.

Hartsuiker, R. J., & Westenberg, C. (2000). Word order priming in written and spoken sentence production. *Cognition*, *75*, 27–39.

Hazenburg, S., & Hulstijn, J. (1996). Defining a minimal receptive second-language vocabulary for non-native university students: An empirical investigation. *Applied Linguistics*, *17*, 145–163.

Hegarty, M., Shah, P., & Miyake, A. (2000). Constrints on using the dual-task methodology to specify the degree of central executive involvement in cognitive tasks. *Memory & Cognition*, *28*, 376–385.

Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Research*, *21*, 303–317.

Henriksen, B., Albrechtsen, D., & Haastrup, K. (2004). The relationship between vocabulary size and reading comprehension in the L2. *Angles on the English-Speaking World*, *4*, 129–140.

Heredia, R. R., & Brown, J. M. (2004). Bilingual memory. In T. K. Bhatia & W. C. Ritchie (Eds.), *The handbook of bilingualism* (pp. 225–249). Wiley-Blackwell.

Heredia, R. R., & Stewart, M. T. (2002). On-line methods in bilingual spoken language research. In R. R. Heredia & J. Altarriba (Eds.), *Bilingual sentence* *processing* (pp. 7–28). Elsevier.

Hermans, D., Bongaerts, T., De Bot, K., & Schereuder, R. (1998). Producing words in a foreign language: Can speakers prevent interference from their first language? *Bilingualism: Language and Cognition*, *1*, 213–229.

Hernandez, A., Li, P., & MacWhinney, B. (2005). The emergence of competing modules in bilingualism. *Trends in Cognitive Sciences*, *9*, 220–225.

Hernandez, A. E. (2009). Language switching in the bilingual brain: What's next? *Brain and Language*, *109*, 133–140.

Hernandez, A. E., Bates, E. A., & Avila, L. X. (1996). Processing across the language boundary: A cross-modal priming study of spanish-english bilinguals. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *22*, 846–864.

Hernandez, A. E., & Meschyan, G. (2006). Executive function is necessary to enhance lexical processing in a less proficient L2: Evidence from fMRI during picture naming. *Bilingualism: Language and Cognition*, *9*, 177–188.

Hewson, L., & Martin, J. (1991). *Redefining translation: The variational approach.* London: Routledge.

Hitch, G. J., & Baddeley, A. D. (1976). Verbal reasoning and working memory. *Quarterly Journal of Experimental Psychology*, *28*, 603–621.

Holmes, J. S. (1987). The name and nature of translation studies. *Indian Journal of Applied Linguistics.*, *13*, 9–24.

Hönig, H., & Kußmaul, P. (1982). *Strategie der übersetzung: ein lehr-und arbeitsbuch.* Gunter: Narr Verlag.

Hoosain, R. (1984). Experiments on digit spans in the Chinese and English languages. In H. S. R. K. Hoosain & R. (Eds.), *Psychological studies of the Chinese Language* (pp. 23–38). Hong Kong: The Chinese Language Society of Hong Kong.

Hoshino, N., Dussias, P. E., & Kroll, J. F. (2010). Processing subject-verb agreement in a second language depends on proficiency. *Bilingualism: Language and Cognition*, *13*, 87–98.

Hoshino, N., & Kroll, J. F. (2008). Cognate effects in picture naming: Does cross-language activation survive a change of script? *Cognition*, *106*, 501–511.

Hromosová, A. (1972). A study of memory in interpreting. *Acta Universitatis*, *17*.

Hsiao, F., & Gibson, E. (2003). Processing relative clauses in chinese. *Cognition*, *90*, 3–27.

Hyltenstam, K. (1992). Non-native features of near-native speakers: On the ultimate attainment of childhood l2 learners. In R. Harris (Ed.), *Cognitive processing in bilinguals.* (pp. 351–368). North-Holland.

Hyltenstam, K., & Abrahamsson, N. (2000). Who can become native-like in a second language? All, some or none? On the maturational constraints controversy in second language acquisition. *Studia Linguistica*, *54*, 150–166.

Hyltenstam, K., & Abrahamsson, N. (2003). Maturational constraints in SLA. In C. I. Doughty & M. H. Long (Eds.), *The Hanbook of Second Language Acquisition.* (pp. 539–588). Blackwell.

Ibáñez, A., Van Hell, J., Macizo, P., Witteman, M. J., & Bajo, M. T. (2008). Is reading for translation different to reading for repetition? An electrophysiological study with dutch/english bilinguals. In *A presentation in the International Conference on Models of Interaction in Bilinguals, 24-26 oct 2008.* Bangor: The University of Wales.

Isham, W. (1994). Memory for sentence form after simultaneous interpretation: Evidence both for and against deverbalization. In S. L. Moser-Mercer & Barbara (Eds.), *Bridging the gap: Empirical research in simultaneous interpretation.* (pp. 191–211). Amsterdam: John Benjamins.

Jacquemot, C., & Scott, S. K. (2006). What is the relationship between phonological short-term memory and speech processing? *Trends in Cognitive Sciences*, *10*, 480–486.

Jakobovits, L. A., & Lambert, W. E. (1961). Semantic satiation among bilinguals. *Journal of Experimental Psychology*, *62*, 576–582.

Jakobson, R. (1966). On linguistic aspects of translation. In R. A. Brower (Ed.), *On translation* (pp. 232–239). New York: Oxford University Press.

Jakubowicz, C., & Nash, L. (2001). Functional categories and syntactic operations in (ab) normal language acquisition. *Brain and language*, *77*, 321–339.

Janis, M. (2002). From the A language to the B language and from the B language to the A language: What is the difference? In *Perspectives on interpreting.* (pp. 53–64). Forli: Biblioteca della Scuola Superiore di Lingue Moderne per Interpreti e Traduttori Forli.

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

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word-frequency effects in speech production - retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology-Learning Memory and Cognition*, *20*(4), 824-843. (Nv233 Times Cited:299 Cited References Count:72)

Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics*, *25*, 603–634.

Jiang, N., & Foster, K. I. (2001). Cross-language priming asymmetires in lexical decision and episodic recognition. *Journal of Memory and Language*, *44*, 32–51.

Jin, Y. (2008). Translation takes place before utterance in a reading for translation task: Evidence from a self-paced reading study. In *Programme of the 4th Psycholinguistic Postgraduate Conference, Department of Psychology, the University of Edinburgh, 2-3 June.*

Johnson, J., Shenkman, K., Newport, E., & Medin, D. (1996). Indeterminacy in the grammar of adult language learners. *Journal of Memory and Language*, *35*(3), 335–352.

Johnson, J. S., & Newport, E. L. (1991). Critical period effects on universal properties of language : The status of subjacency in the acquisition of a second language. *Cognition*, *39*, 215–258.

Jou, J., & Harris, R. J. (1992). The effect of divided attention on speech production. *Bulletin of the Psychonomic Society*, *30*, 301–304.

Just, M., & Carpenter, P. (1992). A capaciity theory of comprehension: individual differences in working memory. *Psychological Review*, *99*, 122–149.

Kade, O. (1968). Kommunikationswissenschaftliche probleme der translation. In *Grundfragen der übersetzungswissenschaft.* Vorlag Enzyklopädie.

Kahnemann, D., & Treisman, A. (1984). Changing views of attention and automaticity. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of attention.* (p. 29-61). Academic Press.

Kaur, S., & Mills, R. (1993). Children as interpreters. In R. Mills & J. Mills (Eds.), *Bilingualism in the primary school.* (pp. 113–125). Routledge.

Kaya, B. (2007, January). *The role of bilingualism in translation activity.* Retrieved from `http://accurapid.com/Journal/39bilingual.htm`

Kempen, G., & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science.*, *11*, 201–258.

Kemper, S. (1987). Syntactic complexity and elderly adults' prose recall. *Experimental Aging Research*, *13*, 47–52.

Kintsch, W. (1974). *The representation of meaning in memory.* Lawrence Erlbaum Associates.

Kintsch, W. (1994). The psychology of discourse processing. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 721–740). Academic Press.

Kintsch, W., Patel, V., & Ericsson, K. (1999). The role of long-term working memory in text comprehension. *Psychologia*, *42*, 186–198.

Kintsch, W., & Van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, *85*, 363–394.

Kirsner, K., Lalor, E., & Hird, K. (1993). The bilingual lexicon: exercise, meaning and morphology. In R. Schreuder & B. Weltens (Eds.), *The bilingual lexicon.* (pp. 215–248). Amsterdam: John Benjamins.

Kiyoshi, K. (2008). 日本における翻研究の展望－理論教育 (Overview of interpreting research in Japan: Theory, Practice and Education). In *The abstract and program of the JAIS 9th annual conference, September 2008, Tokyo: Kyoto university.*

Knudsen, E. I. (2004). Sensitive periods in the development of the brain and behavior. *Journal of Cognitive Neuropsychology*, *16*, 1412–1425.

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

Kormos, J. (2006). *Speech production and second language acquisition.* Mahwah: Erlbaum.

Kotz, S. A. (2009). A critical review of erp and fmri evidence on L2 syntactic processing. *Brain and Language*, *109*, 68–74.

Kroll, J. F. (1993). Accessing conceptual representations for words in a second language. In R. Schreuder & B. Weltens (Eds.), *The bilingual lexicon.* (pp. 53–82). John Benjamins.

Kroll, J. F., & De Groot, A. M. B. (1997). Lexical and conceptual memory in the bilingual: Mapping form to meaning in two languages. In A. M. B. D. Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives.* (pp. 169–199). Lawrence Erlbaum Associates.

Kroll, J. F., & Dijkstra, T. (2005). The bilingual lexicon. In *The Oxford Handbook of Applied Linguistics* (pp. 301–324). Oxford University Press.

Kroll, J. F., Michael, F., Tokowicz, N., & Dufour, R. (2002). The development of lexical fluency in a second language. *Second Language Research*, *18*, 137–171.

Kroll, J. F., & Stewart, E. (1994). Category interference in translation and picture naming: Evidence for asymmetric connections between bilingual memory representations. *Journal of Memory and Language*, *33*, 149–174.

Kroll, J. F., Sumutka, B. M., & Schwartz, A. I. (2005). A cognitive view of the bilingual lexicon: Reading and speaking words in two languages. *International Journal of Bilingualism*, *9*, 27–48.

Kroll, J. F., & Tokowicz, N. (2001). The development of conceptual representation for words in a second language. In J. Nicol (Ed.), *One mind, two languages: Bilingual language processing.* (pp. 49–71). Blackwell.

Kroll, J. F., & Tokowicz, N. (2005). Models of bilingual representation and processing. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches.* (p. 531-553). Oxford University Press.

Kućera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English.* Providence: Brown University Press.

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neurosciences*, *5*, 831–843.

LaHeij, W. (2005). Selection processes in monolingual and bilingual lexical access. In J. F. Kroll & A. M. B. D. Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 289–307). Oxford University Press.

La Heij, W., De Bruyn, E., Elens, E., Hartsuiker, R., Helha, D., & Van Schelven, L. (1990). Orthographic facilitation and categorical interference in a word-translation variant of the stroop task. *Canadian Journal of Psychology*, *44*, 76–83.

La Heij, W., Kerling, R., & Velden, E. van der. (1996). Nonverbal context effects in forward and backward translation: Evidence for concept mediation. *Journal of Memory and Language*, *35*, 648-665.

Lambert, S. (1983). *Recognition and recall among conference interpreters.* Unpublished Ph.D. thesis, University of Stirling.

Lambert, W. E. (1963). Psychological approach to the study of language. On second-language learning and bilingualism. *The Modern Language Journal*, *47*, 114–121.

Lambert, W. E., & Fillenbaum, S. (1959). A pilot study of aphasia among bilinguals. *Canadian Journal of Psychology*, *13*, 28–34.

Lambert, W. E., Havelka, J., & Crosby, C. (1958). The influence of language-acquisition contexts on bilingualism. *The Journal of Abnormal and Social Psychology*, *56*, 239–244.

Lardiere, D. (1998). Case and tense in the fossilized state grammar. *Second Language Research*, *14*, 1-26.

Laufer, B. (1992). How much lexis is necessary for reading comprehension? In P. J. L. Arnaud & H. Bejoint (Eds.), *Vocabulary and applied linguistics* (p. 126-132). London: Macmillan.

Laufer, B. (1998). The development of passive and active vocabulary in a second language: Same or Different? *Applied Linguistics*, *19*, 255–271.

Laufer, B., Elder, C., Hill, K., & Congdon, P. (2004). Size and strength: do we need both to measure vocabulary knowledge? *Language Testing*, *21*, 202–226.

Laufer, B., & Nation, P. (2001). Passive vocabulary size and speed of meaning recognition: Are they related? *EUROSLA Yearbook*, *1*, 7–28.

Lee, Y. H. (2003). *Comparison of error frequency in simultaneous interpretation A to B and B to A (Korean-English).* Unpublished (pre-doctoral) thesis, University of Geneva, Ecole de Traduction et Interpretation.

Lenneberg, E. H. (1967). *Biological foundations of language.* Wiley.

Lennon, P. (1984). Retelling a story in english as a second language. In H. W. Dechert, D. Mohle, & M. Raupach (Eds.), *Second language productions.* (pp. 50–68). Gunter: Narr Verlag.

Levelt, W. J. M. (1989). *Speaking: from intention to articulation.* Cambridge: MIT Press.

Levelt, W. J. M. (1992). Accessing words in speech production: stages, processes and representations. *Cognition*, *42*, 1–22.

Levelt, W. J. M., & Kelter, S. (1982). Surface form and memory in question answering. *Cognitive Psycholgy*, *14*, 78–106.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioural and Brain Science*, *22*, 1–38.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*, 1126–1177.

Lewandowsky, S., Oberauer, K., & Brown, G. D. A. (2009). No temporal decay in verbal short-term memory. *Trends in Cognitve Sciences*, *13*, 120–126.

Li, C. N., & Thompson, S. A. (1981). *Mandarin Chinese: A functional reference grammar*. London: University of California Press.

Li, L. (1998). A comparison of word order in english and chinese. *Papers and studies in contrastive linguistics.*, *34*, 153–161.

Lin, C.-J., & Bever, T. G. (2006). Subject preference in the processing of relative clauses in Chinese. In M. S. D. Baumer & D. Montero (Eds.), *Proceedings of the 25th West Coast Conference of Formal Linguistics.* (pp. 254–260).

Linderholm, T., & Broek, P. Van den. (2002). The effects of reading purpose and working memory capacity on the processing of expository text. *Journal of Educational Psychology*, *94*, 778–784.

Lindsley, J. R. (1976). Producing simple utterances: Details of the planning process. *Journal of Psycholinguistic Research*, *5*, 331–354.

Liu, I. M., Wu, J. T., & Chou, T. L. (1996). Encoding operation and transcoding as the major loci of the frequency effect. *Cognition*, *59*, 149–168.

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

Liu, Y., & Perfetti, C. A. (2003). The time course of brain activity in reading English and Chinese: An ERP study of Chinese bilinguals. *Human Brain*

Mapping, *18*, 167–175.

Loebell, H., & Bock, J. K. (2003). Structural priming across languages. *Linguistics*, *41*, 791–824.

Logan, G. D., & Cowan, W. B. (1984). On the ability to inhibit thought and action: A theory of an act of control. *Psychological Review*, *91*, 295–327.

Logie, R. H. (2003). Spatial and visual working memory: A mental workspace. In B. H. Ross & D. Irwin (Eds.), *Cognitive vision: Psychology of learning and motivation* (pp. 37–78). Oxford: Gulf Professional Publishing.

Logie, R. H. (2009, December). *Explore your memory.* Retrieved from `http://www.bbc.co.uk/science/humanbody/mind/surveys/memory/`

Logie, R. H., Della Sala, S., Wynn, V., & Baddeley, A. D. (2000). Visual similarity effects in immediate verbal serial recall. *Quarterly Journal of Experimental Psychology A*, *53*, 626–646.

Logie, R. H., & Duff, S. (2007). Separating processing from storage in working memory operation span. In N. Osaka, R. H. Logie, & M. D'Esposito (Eds.), *The cognitive neuroscience of working memory.* (pp. 119–135). Oxford: Oxford University Press.

Logie, R. H., & Maylor, E. A. (2009). An internet study of prospective memory across adulthood. *Psychology and Aging.*, *24*, 767–774.

Long, M. H. (1990). Maturational constraints on lanuage development. *Studies in Second Language Acquisition*, *12*, 251–268.

Lörscher, W. (1989). Models of the translation process: Claim and reality. *Target: International Journal of Translation Studies*, *1*, 43–68.

Luftig, R. (1982). Normalization in paraphrase and recall: Effects of processing, grammatical article type, and retention interval. *Journal of Psycholinguistic Research*, *11*, 127–140.

MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguitiy. *Cognitive Psychology*, *24*, 56–98.

Macizo, P., & Bajo, M. T. (2004). When translation makes the difference: Sentence procesing in reading and translation. *Psicologica*, *25*, 181–205.

Macizo, P., & Bajo, M. T. (2006). Reading for repetition and reading for translation: Do they involve the same processes? *Cognition*, *99*, 1–34.

Macizo, P., & Bajo, M. T. (2009). Schema activation in translation and reading: A paradoxical effect. *Psicológica*, *30*, 59–89.

MacKay, D. G., & Bowman, R. W. (1969). On producing the meaning in sentences. *American Journal of Psychology.*, *82*, 23-39.

MacKintosh, J. (1985). The kintsch and van dijk model of discourse comprehension and production applied to the interpretation process. *Meta*, *30*, 37–43.

Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spotaneous English speech. *Word*, *15*, 19–44.

MacWhinney, B. (1999). The emergence of language from embodiment. In B. MacWhinney (Ed.), *The emergence of language.* (pp. 213–256). Lawrence Erlbaum.

Malakoff, M., & Hakuta, K. (1991). Translation skill and metalinguistic awareness in bilinguals. In *Langauge processing in bilingual children.* (pp. 141–166). Cambridge: Cambridge University Press.

Marian, V., & Neisser, U. (2000). Language-dependent recall of autobiographical memories. *Journal of Experimental Psychology: General*, *129*, 361–368.

Marian, V., & Spivey, M. (2003). Competing activation in bilingual language procesing: Within- and between-language competition. *Bilingualism: Language and Cognition*, *6*, 97–115.

Marinova-Todd, S. (2003). *Comprehensive analysis of ultimate attainment in adult second language acquisition.* Unpublished doctoral dissertation, Harvard University.

Massaro, D. W. (1975). *Understanding language.* New York: Academic Press.

Matysiak, A. (2001). *Controlled processing in simultaneous interpretation: A study based on daniel gile's effort models.* Unpublished master's thesis, University of Poznan.

Mazurkewich, I., & White, L. (1984). The acquisition of the dative alternation: Unlearning overgeneralizations. *Cognition*, *16*, 261–283.

McClelland, J. L. (1979). Time relations of mental processes - Examination of systems of processes in cascade. *Psychological Review*, *86*, 287–330.

McDonald, J. L. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language*, *55*, 381–401.

McDonough, K. (2006). Interaction and syntactic priming: English l2 speakers' production of dative constructions. *Studies in Second Language Acquisition*, *28*, 179-207.

McKoon, G., Ratcliff, R., Ward, G., & Sproat, R. (1993). Syntactic prominence effect on discourse processes. *Journal of Memory and Language*, *32*, 593–607.

McLaughlin, B. (1978). *Second language acquisition in childhood.* Erlbaum.

McNulty, J. A. (1966). A partial learning model of recognition memory. *Canadian Journal of Psychology*, *20*, 302–315.

Meara, P. M. (1996). The dimensions of lexical competence. In G. Brown, K. Malmkjaer, & J. Williams (Eds.), *Performance and competence in second language acquisition* (pp. 35–53). Cambridge: Cambridge University Press.

Meisel, J. M. (2008). Child second language acquisition or successive first language acquisition? In B. Haznedar & E. Gavruseva (Eds.), *Current Trends in Child Second Language Acquisition: A Generative Perspective.* (pp. 55–82). Amsderdam: John Benjamins.

Melka, F. (1997). Receptive vs. productive aspects of vocabulary. In N. Schmitt & M. McCarthy (Eds.), *Vocabulary: Description, Acquisition and Pedagogy* (pp. 84–102). Cambridge: Cambridge Universtiy Press.

Meyer, B. J. F. (1975). *The organization of prose and its effect on memory.* North-Holland.

Mezynski, K. (1983). Issues concerning the acquisition of knowledge: Effects of vocabulary training on reading comprehension. *Review of Educational Research*, *53*, 253–279.

Michael, E. B., Dijkstra, T., & Kroll, J. F. (2002). *Individual differences in the degree of language nonselectivity in fluent bilinguals.* (Poster presented at the annual meeting of the Psychonomic Society, 21–24 November, Kansas City)

Michael, E. B., & Gollan, T. H. (2005). Being and becoming bilingual – individual differences and consequences for language production. In J. F. Kroll & A. M. B. D. Groot (Eds.), *Handbook of bilingualism – psycholinguistic approaches.* (p. 389-410). Oxford University Press.

Miller, J. R., & Kintsch, W. (1980). Readability and recall of short prose passages: A theoretical analysis. *Journal of Experimental Psychology: Human, Learning, and Memory*, *6*, 335–354.

Milton, J., Wade, J., & Hopkins, N. (in press). Aural word recognition and oral competence in a foreign language. In R. Chacón-Beltrán, C. Abello-Contesse, M. Torreblanca-López, & M. López-Jiménez (Eds.), *Further insights into non-native vocabulary teaching and learning.*

Miyake, A., Carpenter, P. A., & Just, M. A. (1994). A capacity approach to syntactic comprehension disorders: Making normal adults perform like aphasic patients. *Cognitive Neuropsychology*, *11*, 671–717.

Miyake, A., & Friedman, N. P. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In *Foreign language learning: Psycholinguistic studies on training and retention.* (pp. 339–364). Mahwah, NJ: Erlbaum.

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

Miyake, A., & Shah, P. (Eds.). (1999). *Models of working memory: An introduction.* Cambridge: Cambridge University Press.

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

Moser-Mercer, B. (1978). Simultneous interpretation: A hypothetical model and its practical application. In D. Gerver & H. W. Sinaiko (Eds.), *Language interpretation and communication* (pp. 353–368). Plenum Press.

Moser-Mercer, B. (2002). Process models in simultaneous interpretation. In F. Pöchhacker & M. Shlesinger (Eds.), *The interpreting studies reader.* (pp. 120–129). Routledge.

Murphy, G. L., & Shapiro, A. M. (1994). Forgetting of verbatim information in discourse. *Memory & Cognition*, *22*, 85–94.

Nation, I. S. P. (1990). *Teaching and learning vocabulary.* Newbury House.

Nation, I. S. P. (2001). *Learning vocabulary in another language.* Cambridge University Press.

Nation, I. S. P. (2006). How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review*, *63*, 59–82.

Naveh-Benjamin, M., Guez, J., & Marom, M. (2003). The effects of divided attention at encoding on item and associative memory. *Memory & Cognition*, *31*, 1021–1035.

Navon, D., & Gopher, D. (1979). On the economy of the human-processing system. *Psychological Review.*, *86*, 214–255.

Neely, J. H. (1977). Semantic priming and retrieval from lexical memory: Roles of inhibitionless spreading activation and leimited-capacity attention. *Journal of Experimental Psychology: General*, *106*, 226–254.

Neubert, A. (2000). Competence in language, in languages, and in translation. In C. Schaffner & B. Adab (Eds.), *Developing Translation Competence.* (pp.

3–18). Amsterdam: John Benjamins.

Nicol, J., & Greth, D. (2003). Production of subject-verb agreement in spanish as a second language. *Experimental Psychology*, *50*, 196–203.

Nicol, J., Teller, M., & Greth, D. (2001). Production of verb agreement in monolingual, bilingual and second-language speakers. In J. Nicol (Ed.), *One mind, two languages: bilingual language processing.* (pp. 117–133). Wiley-Blackwell.

Nida, E. A. (1969). Science of translation. *Language*, *45*, 483–498.

Nida, E. A. (1991). Theories of translation. *TTR: Traduction, Terminologie, Rédaction*, *4*, 19–32.

Nida, E. A., & Taber, C. R. (2003). *The theory and practice of translation.* Leiden: Brill.

Odlin, T. (2003). Cross-linguistic influence. In C. J. Doughty & M. H. Long (Eds.), *Handbook of Second Language Acquisition.* (pp. 436–486). Blackwell.

Oomen, C. C. E., & Postma, A. (2001). Effects of divided attention on the production of filled pauses and repetitions. *Journal of Speech, Language and Hearing Research*, *44*, 997.

Orellana, M. F., Dorner, L., & Pulido, L. (2003). Accessing assets: Immigrant youth's work as family translators or "paraphrasers". *Social Problems*, *50*, 505–524.

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*, 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* (p. 61-72). Turku: University of Turku, Centre for Translation and Interpreting.

Padilla, P., Bajo, M. T., & Padilla, F. (1999). Proposal for a cognitive theory of translation and interpreting: A methodology for future empirical research.

The Interpreters' Newsletter, *9*, 61–78.

Pallier, C., Dehaene, S., Poline, J. B., LeBihan, D., Argenti, A. M., Dupoux, E., et al. (2003). Brain imaging of language plasticity in adopted adults: can a second language replace the first? *Cerebral cortex*, *13*, 155.

Panico, J., & Healey, E. C. (2008). The influence of text type, topic familiarity, and stuttering frequency on listener recall, comprehension, and mental effort. *Journal of Speech, Language, and Hearing Research*, *52*, 534–546.

Papagno, C., Valentine, T., & Baddeley, A. D. (1991). Phonological short-term-memory and foreign-language vocabulary learning. *Journal of Memory and Language*, *30*, 331–347.

Paradis, M. (1987). *The assessment of bilingual aphasia.* Hillsdale: Erlbaum.

Paradis, M. (1994a). Neurolinguistic aspects of implicit and explicit memory: Implications for bilingualism. In N. Ellis (Ed.), *Implicit and explicit language learning.* London: Academic Press.

Paradis, M. (1994b). Toward a neurolinguistic theory of simultaneous translation: The framework. *International Journal of Psycholinguistics*, *10*, 319-335.

Paradis, M. (2009). *Declarative and procedural determinants of second languages.* Amsterdam: John Benjamins.

Pavlenko, A. (2009). Conceptual representation in the bilingual lexicon and second language vocabulary learning. In A. Pavlenko (Ed.), *The bilingual mental lexicon: Interdisciplinary approaches* (pp. 125–160). Bristol: Multilingual Matters.

Perea, M., Dunabeitia, J. A., & Carreiras, M. (2008). Masked associative/semantic priming effects across languaages with highly proficient bilinguals. *Journal of Memory and Language*, *58*, 916–930.

Phillips, N. A., Klein, D., Mercier, J., & deBoysson, C. (2006). ERP measures of auditory word repetition and translation priming in bilinguals. *Brain Research*, *1125*, 116–131.

Pickering, M. J., & Branigan, H. P. (1998). The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language*, *39*, 633–651.

Pickering, M. J., & Branigan, H. P. (1999). Syntactic priming in language production. *Trends in Cognitive Sciences*, *3*, 136–141.

Pickering, M. J., Branigan, H. P., & McLean, J. F. (2002). Constituent structure is formulated in one stage. *Journal of Memory and Language*, *46*, 586–605.

Pickering, M. J., & Ferreira, V. (2008). Structural priming: a critical review. *Psychological Bulletin*, *134*, 427–459.

Pienemann, M. (2005). *Cross-linguistic aspects of processability theory.* John Benjamins Publishing Company.

Pöchhacker, F. (2004a). From operation to action: process-orientation in interpreting studies. *Meta*, *50*, 682–695.

Pöchhacker, F. (2004b). *Introducing interpreting studies.* Routledge.

Pöchhacker, F. (2008). The turns of interpreting studies. In *Efforts and models in interpreting and translation research.* John Benjamins.

Posner, M., & Boies, S. (1971). Components of attention. *Psychological Review*, *78*, 391–408.

Potter, M. C., & Lombardi, L. (1998). Syntactic priming in immediate recall of sentences. *Journal of Memory and Language*, *38*, 265–282.

Poulisse, N. (1997). Language production in bilinguals. In A. M. B. D. Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives.* (pp. 201–224). Lawrence Erlbaum Associates.

Poulisse, N. (1999). *Slips of the tongue: Speech errors in first and second language production.* Amsterdam: John Benjamins.

Poulisse, N., & Bongaerts, T. (1994). 1st language use in 2nd-language production. *Applied Linguistics*, *15*, 36–57.

Power, M. J. (1986). A technique for measuring processing load during speech production. *Journal of Psycholinguistic Research*, *15*, 371–382.

Presas, M. (2000). Bilingual competence and translation competence. In C. Schäffner & B. Adab (Eds.), *Developing Translation Competence.* (pp. 19–32). Amsterdam: John Benjamins.

Price, C. J., Green, D. W., & Studnitz, R. V. (1999). A functional imaging study of translation and language switching. *Brain*, *122*, 2221–2235.

Prior, A., & MacWhinney, B. (2007). Translation norms for english and spanish: The role of lexical variables, word class, and L2 proficiency in negotiating translation ambiguity. *Behavior Research Methods*, *39*, 1029–1038.

Prodeau, M. (2005). Gender and number in french L2: Can we find out more about the constraints on production in L2? In J.-M. Dewaele (Ed.), *Focus on french as a foreign language: Multidisciplinary approaches* (pp. 135–163). Channel View Publications.

Prévost, P., & White, L. (2000). Accounting for morphological variation in second language acquisition: truncation or missing inflection? In *The acquisition of syntax: studies in comparative developmental linguistics.* (p. 202-235). Longman.

Qian, D. D. (2002). Investigating the relationship between vocabulary knowledge and academic reading performance: An assessment perspective. *Language Learning*, *3*, 513-536.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*, 93–134.

Radvansky, G. A., & Copeland, D. E. (2004). Working memory span and situation model processing. *American Journal of Psychology*, *117*, 191–213.

Raney, G. E., Obeidallah, S. M., & Miura, T. K. (2002). Text comprehension in bilinguals: Integrating perspectives on language representation and text processing. In R. Heredia & J. Altarriba (Eds.), *Bilingual sentence processing.* (pp. 165–183). North-Holland.

Rappaport, M., Laughren, M., & Levin, B. (1987). *Levels of lexical representation. Lexicon project working papers.* MIT.

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: effects of word frequency, verb complexity, and lexical ambiguity. *Mem Cognit*, *14*, 191–201.

Read, J. (2000). *Assessing vocabulary.* Cambridge: Cambridge University Press.

Remijsen, B. (2009, November). *Bert Remijsen's Praat scripts.* Retrieved from `http://www.ling.ed.ac.uk/~bert/praatscripts.html`

R. H., L., G., C., S., D. S., & A. D., B. (2004). Is there a specific executive capacity for dual task coordination? Evidence from Alzheimer's Disease. *Neuropsychology*, *18*, 504–513.

Riccardi, A. (2002). *Translation studies: Perspectives on an emerging discipline.* Cambridge: Cambridge University Press.

Robbins, T., Anderson, E., Barker, D., Bradley, A., Fearneyhough, C., Henson, R., et al. (1996). Working memory in chess. *Memory and Cognition*, *24*, 83–93.

Roberts, R., & Gibson, E. (2002). Individual differences in sentence memory. *Journal of Psycholinguitic Research*, *31*, 573–598.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*, 107–142.

Roelofs, A. (1997). The weaver model of work-form encoding in speech production. *Cognition*, *64*, 249–284.

Roelofs, A. (2003). Shared phonological encoding proceses and represnetations of languages in bilingual speakers. *Language and Cognitive Processes*, *18*, 175–204.

Roelofs, A., & Verhoef, K. (2006). Modeling the control of phonological encoding in bilingual speakers. *Bilingualism: Language and Cognition*, *9*, 167-176.

Rohde, D. (2009, November). *Linger: A flexible platform for language processing experiments, version 2.94.* Retrieved from `http://tedlab.mit.edu/~dr/Linger/`

Rubin, D. C. (1978). A unit analysis of prose memory. *Journal of Verbal Learning and Verbal Behavior*, *17*, 599–620.

Rubin, J. (1994). A review of second language listening comprehension research. *The Modern Language Journal*, *78*, 199–221.

Ruchkin, D. S., Grafman, J., Cameron, K., & Berndt, R. S. (2003). Working memory: Unemployed but still doing day labor. *Behavioral and Brain Sciences*, *26*, 760-777.

Ruiz, C., Paredes, N., Macizo, P., & Bajo, M. T. (2008). Activation of lexical and syntactic target language properties in translation. *Acta Psychologica*, *128*, 490-500.

Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics*, *2*, 437–442.

Saito, S., Logie, R. H., Morita, A., & Law, A. (2008). Visual and phonological similarity effects in verbal immediate serial recall: A test with kanji materials. *Journal of Memory and Language*, *59*, 1–17.

Sanz, J. (1931). Le travail et les aptitudes des interprètes parlementaires (The task and the skills of parliamentary interpreters). *Anals d'Orientació Professional*, *4*, 303–318.

Schachter, J. (1989). Testing a proposed universal. In S. M. Gass & J. Schachter (Eds.), *Linguistic perspectives on second langauge acquisition.* (pp. 73–88). Cambridge: Cambridge University Press.

Schoonbaert, S., Duyck, W., Brysbaert, M., & Hartsuiker, R. J. (2009). Semantic and translation priming from a first language to a second and back: Making sense of the findings. *Memory & Cognition*, *37*, 569–586.

Schoonbaert, S., Hartsuiker, R. J., & Pickering, M. J. (2007). The representation of lexical and syntactic information in bilinguals: Evidence from syntactic priming. *Journal of Memory and Language*, *56*, 153–171.

Schreuder, R., & Weltens, B. (1993). The bilingual lexicon: An overview. In R. Schreuder & B. Weltens (Eds.), *The Bilingual Lexicon* (pp. 1–11). John Benjamins.

Schwartz, A. I., & Kroll, J. F. (2006). Bilingual lexical activation in sentence context. *Journal of Memory and Language*, *55*, 197–212.

Segalowitz, N. (1986). Skilled reading in the second language. In J. Vaid (Ed.), *Language processing in bilinguals: Psycholinguistic and neuropsychological perspectives.* (pp. 3–19). Erlbaum.

Segalowitz, N. (2003). Automaticity and second languages. In C. J. Doughty & M. H. Long (Eds.), *Handbook of second language acquisition.* (pp. 383–408). Blackwell.

Segalowitz, N., & Hebert, M. (1990). Phonological recoing in the first and second language reading of skilled bilinguals. *Language Learning*, *40*, 503–538.

Segalowitz, N., & Hulstijn, J. (2005). Automaticity in bilingualism and second language learning. In *Handbook of bilingualism: Psycholinguistic approaches.* Oxford University Press.

Seidenberg, M. S. (1985). The time course of phonological code activation in two writing systems. *Cognition*, *19*, 1–30.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, *96*, 523–568.

Seleskovitch, D. (1976). Interpertation: A psychological approach to translating. In R. W. Brislin (Ed.), *Translation: Application and Research.* (pp. 64–76). New York: Gardner.

Seleskovitch, D., & Lederer, M. (1992). *Interpréter pour traduire (Chinese version "Introduction to Interpreting and Translating" translated by Sun, H-S).*

Beijing Language and Culture University Press.

Service, E. (1992). Phonology, working memory, and foreign language learning. *The Quarterly Journal of Experimental Psychololgy.*, *45A*, 21-50.

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

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

Shallice, T., & Warrington, E. K. (1974). The dissociation between short term retention of meaningful sounds and verbal material. *Neuropsychologia*, *12*, 553–555.

Shlesinger, M. (2000). *Strategic allocation of working memory and other attentional resources.* Unpublished doctoral dissertation, Bar-Ilan University.

Silverberg, S., & Samuel, A. G. (2004). The effect of age of second language acquisition on the representation and processing of second language words. *Journal of Memory and Language*, *51*, 381–398.

Sjørup, A. C., Jensen, K. T. H., & Balling, L. W. (2009). Syntactic processing in translation from L1 to L2. In *The program of the 'Eye-to-IT Conference on Translation Processes', 28-29 April 2009, Copenhagen, Denmark.*

Smith, M., & Wheeldon, L. (2001). Syntactic priming in spoken sentence production – an online study. *Cognition*, *78*, 123–164.

Sorace, A. (2003). Near-nativeness. In M. H. L. C. Doughty (Ed.), *The handbook of second language acquisition.* (pp. 130–151). Wiley-Blackwell.

Sorace, A. (2005). Selective optionality in language development. In K. P. C. L. E. A. Cornips (Ed.), *Syntax and variation: reconciling the biological and the social* (pp. 55–80). John Benjamins.

Staehr, L. S. (2008). Vocabulary size and the skills of listening, reading and writing. *Language Learning*, *36*, 139–152.

Stein, D. (1980). *Theoretische grundlagen der übersetzungswissenschaft.* Gunter: Narr Verlag.

Sumiya, H., & Healy, A. F. (2008). The stroop effect in English-Japanese bilinguals: The effect of phonological similarity. *Experimental Psychology*, *55*, 93–101.

Sun, C. (2006). *Chinese: A linguistic introduction.* Cambridge: Cambridge University Press.

Swain, M. (1998). Focus on form through conscious reflection. In C. Doughty & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 64–84). Cambridge University Press.

Synovitz, R. (2008, September). *Mistakes by Afghan translators endanger lives hamper antiterrorism effort.* Retrieved from http://www.rferl.org/content/Mistakes_By_Translators_Hamper_Afghan_Antiterrorism_Campaign/1195783.html

Tannenbaum, P. H., & Williams, F. (1968). Generation of active and passive sentences as a function of subject and object focus. *Journal of Verbal Learning and Verbal Behavior.*, *7*, 246–250.

Thelen, M. (2005). Translating into english as a non-native language: The Dutch connection. In G. M. Anderman & M. Rogers (Eds.), *In and out of english: for better, for worse?* (pp. 242–255). Multilingual Matters.

Thierry, G., & Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 12530–12535.

Thorn, A. S. C., Gathercole, S. E., & Frankish, C. R. (2002). Language familiarity effects in short-term memory: The role of output delay and long-term knowledge. *Quarterly Journal of Experimental Psychology A*, *55A*, 1363–1383.

Titone, D., Prentice, K. J., & Wingfield, A. (2000). Resource allocation during spoken discourse processing: Effects of age and passage difficulty as revealed by self-paced listening. *Memory & cognition*, *28*, 1029–1040.

Tokowicz, N., & Kroll, J. F. (2007). Number of meanings and concreteness: Consequences of ambiguity within and across languages. *Language and Cognitive Processes*, *22*, 727–779.

Tommola, J., & Helevä, M. (1998). Language direction and source text complexity. In L. Bowker (Ed.), *Unity in diversity? Current trends in translation studies* (pp. 177–186). St. Jerome.

Toppelberg, C. O. (1997). Biopsychosocial model. Author's replies. *The American Journal of Psychiatry.*, *154*, 1328–1329.

Toury, G. (1995). *Descriptive translation studies and beyond.* John Benjamins.

Towse, J. N., & Hitch, G. J. (1995). Is there a relationship between task demand and storage space in tests of working memory capacity? *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, *48A*, 108–124.

Trueswell, J. C., Michael, K., Tanenhaus, K., & Garnsey, S. M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, *33*, 285–318.

Turner, A., & Greene, E. (1977). *The construction and use of a propositional text base.* Boulder: Institute for the Study of Intellectual Behavior, University of Colorado.

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

Ullman, M. T. (2001). The neural basis of lexicon and grammar in first and second language: the declarative/procedural model. *Bilingualism: Language and Cognition*, *4*(02), 105-122 M3 - 10.1017/S1366728901000220. (1366-7289)

Ullman, M. T. (2004). Contributions of memory circuits to language: the declarative/procedural model. *Cognition*, *92*(1-2), 231-270. (0010-0277 doi: DOI: 10.1016/j.cognition.2003.10.008)

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*, 498–505.

Vallar, G., & Baddeley, A. D. (1987). Phonological short-term store and sentence processing. *Cognitive Neuropsychology*, *4*, 417–438.

Van Assche, E., Duyck, W., Hartsuiker, R. J., & Diependaele, K. (2009). Does bilingualism change native-language reading? Cognate effects in a sentence context. *Psychological Science*, *20*, 923–927.

Van Dijk, T. A., & Kintsch, W. (1983). *Strategies of discourse comprehension.* Academic Press.

Van Hell, J. G., & De Groot, A. M. B. (2008). Sentence context modulates visual word recognition and translation in bilinguals. *Acta Psychologica*, *128*, 431–451.

Van Hoof, H. (1962). *Théorie et pratique de la traduction.* München: Hueber.

Vasos, H. (1983). *Semantic processing in bilinguals.* Unpublished doctoral dissertation, Concordia Univeristy, Montreal.

Verhoef, K., Roelofs, A., & Chwilla, D. J. (2009). Role of inhibition in language switching: Evidence from event-related brain potentials in overt picture naming. *Cognition*, *110*, 84–99.

Vitevitch, M. S. (2002). The influence of phonological similarity neighborhoods on speech production. *Journal Of Experimental Psychology: Learning, Memory and Cognition*, *28*, 735–747.

Voga, M., & Grainger, J. (2007). Cognate status and cross-script translation priming. *Memory & Cognition*, *35*, 938–952.

Von Stutterheim, C., & Nüse, R. (2003). Processes of conceptualization in language production: Language-specific perspectives and event construal. *Linguistics*, *41*, 851–881.

Voss, B. (1984). *Slips of the ear: investigations into the speech perception behaviour of german speakers of english.* John Benjamins.

Waters, G. S., & Caplan, D. (1996a). The capacity theory of sentence comprehension: Critique of just and carpenter (1992). *Psychological Review*, *103*, 761–772.

Waters, G. S., & Caplan, D. (1996b). The measurement of verbal working memory capacity and its relation to reading comprehension. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology*, *49*(1), 51-79. (Uf604 Times Cited:84 Cited References Count:26)

Waters, G. S., & Caplan, D. (2002). Working memory and online syntactic processing in Alzheimer's Disease: Studies with auditory moving window presentation. *Journal of Gerontology: Psychological Sciences and Social Sciences*, *57*, 298–311.

Waters, G. S., Caplan, D., & Yampolsky, S. (2003). On-line syntactic processing under concurrent memory load. *Psychonomic Bulletin & Review*, *10*, 88–95.

Webb, S. (2008). Receptive and productive vocabulary sizes of L2 learners. *Studies in Second Language Acquisition*, *30*, 79–95.

Weber-Fox, C., & Neville, H. J. (2001). Sensitive periods differntiate processing of open- and closed-class words. *Journal of Speech, Language, and Hearing Research*, *44*, 1338–1353.

Weinreich, U. (1953). *Languages in contact: Findings and problems.* New York: Linguistic Circle of New York.

Wiese, R. (1984). Language production in foreign and native languages: Same or different? In H. W. Dechert, D. Mohle, & M. Raupach (Eds.), *Second language productions* (pp. 11–25). Gunter: Narr Verlag.

Wikipedia. (2009, October). *Tagalog language.* Retrieved from `http://en.wikipedia.org/wiki/Tagaloglanguage`

Williams, J. (1992). Planning, discourse marking, and the comprehensibility of international teaching assistant. *TESOL Quarterly*, *26*, 693–711.

Williams, S. (1995). Research on bilingualism and its relevance for interpreting. *Journal of Linguistics*, *15*, 143–154.

Wilson, B., & Baddeley, A. D. (1988). Semantic, episodic, and autobiographical memory in a postmeningitic amnesic patient. *Brain & Cognition*, *8*, 31–46.

Wilson, D., & Sperber, D. (1979). Ordered entailments: An alternative to presuppositional theories. *Syntax and Semantics*, *11*, 299–323.

Wingfield, A., & Stine, E. L. (1986). Organizational strategies in immediate recall of rapid speech by young and elderly adults. *Experimental Aging Research*, *12*, 79–83.

Yang, I. L. (2006). On the issue of vocabulary size in english teaching in Taiwan. *Journal of the National Institute for Compilation and Translation*, 35–44.

Yip, P.-C., & Rimmington, D. (2004). *Chinese: A comprehensive grammar.* London: Routledge.

Yuan, F., & Ellis, R. (2003). The effects of pre-task planning and on-line planning on fluency, complexity and accuracy in L2 monologic oral production. *Applied Linguistics*, *24*, 1–27.

Zwaan, R. A., Ericsson, K. A., Lally, C., & Hill, L. (1998). Situation model construction during translation. *Unpublished essay*.

Zwaan, R. A., & Singer, M. (2003). Text comprehension. In A. C. Graesser, M. A. Gernsbacher, & S. R. Goldman (Eds.), *Handbook of discourse processes* (pp. 83–122). Lawrence Erlbaum Associates.

Appendix A

Tables for the Experiments

Experiment 1

Table A.1: Proportions and standard deviations of memory performance in discourse interpreting among post-graduate students

Direction	English-Chinese	English-Chinese	Chinese-English	Chinese-English
Congruency	Congruent	Incongruent	Congruent	Incongruent
Mean(SD)	36.4(10)	41.6(11.8)	38.2(10.6)	58.2(16)

Experiment 2

Table A.2: Proportions and standard deviations of memory performance in discourse interpreting among interpreting students

Direction	English-Chinese	English-Chinese	Chinese-English	Chinese-English
Congruency	Congruent	Incongruent	Congruent	Incongruent
Mean(SD)	46.5(15.3)	46.5(14.3)	51(8.4)	65.6(8.3)

Experiment 3

Table A.3: Proportions and standard deviations of memory performance in discourse paraphrasing and verbatim recall

Instruction	Verbatim	Verbatim	Paraphrase	Paraphrase
Type	Congruent	Incongruent	Congruent	Incongruent
Mean(SD)	44.6(11.9)	66.2(9.9)	54.3(13)	55(12.1)

Experiment 5

Table A.4: Mean reading time in four conditions

Recall Method	Repeat	Repeat	Translate	Translate
Word order	Congruent	Incongruent	Congruent	Incongruent
Mean(SD)	857(335)	915(464)	1479(863)	1977(1467)

Table A.5: Mean percentage of correctly reproduced propositions in four conditions

Recall Method	Repeat	Repeat	Translate	Translate
Word order	Congruent	Incongruent	Congruent	Incongruent
Mean(SD)	95.8(3.1)	95.9(4.3)	89.3(5.8)	89(9.5)

Experiment 6

Table A.6: Mean Reading Time of critical word in all conditions

Preload	No-load	No-load	5-digt	5-digit	7-digit	7-digit
Word order	Congruent	Incongruent	Congruent	Incongruent	Congruent	Incongruent
Mean(SD)	1106(552)	1408(1178)	901(482)	1084(873)	905(511)	983(714)

Experiment 7

Table A.7: Mean Reading Time of critical word in all conditions

Preload	No-load	No-load	5-digt	5-digit
Word order	Congruent	Incongruent	Congruent	Incongruent
Mean(SD)	1471(1067)	1644(1265)	1277(729)	1330(1283)

Speech Rate Analysis

Table A.8: Mean speech rate in different conditions of sentence interpreting

Preload	No-load	No-load	5-digt	5-digit	7-digit	7-digit
Word order	Congruent	Incongruent	Congruent	Incongruent	Congruent	Incongruent
Mean(SD)	2.16(.38)	2.33(.33)	2.25(.44)	2.27(.45)	2.17(.41)	2.34(.41)

Experiment 8

Table A.9: Mean reading time in four conditions. Congruent=Non-Ba sentence; Incongruent= Ba-sentences

Recall Method	Repeat	Repeat	Translate	Translate
Word order	Congruent	Incongruent	Congruent	Incongruent
Mean(SD)	734(288)	844(464)	1640(1082)	1958(1036)

Appendix B

Material for Experiment 1-4

Chinese Congruent Discourses

1. 疼痛有著強烈的心理成份。患者心裡想些什麼能決定痛覺的強度。轉移患者的注意力，像是讓他們聽音樂，可以協助降低疼痛感。人類注意力讓我們在選擇某些資訊來處理時，會忽略其他的。
2. 我們已經知道限制一隻動物攝取的熱量可以延長牠的壽命。儘管這個方法被發現是在七十年前，它仍然是唯一絕對有效的方法。熱量限制法一般是減少動物飲食量的百分之30到40。
3. 85 90%的口臭來源是口腔。如果一個健康的人有口臭，主要來源是是舌頭，而不是牙齒。除了口腔以外，另一個主要來源則是鼻子。這種口臭有不同的特質。
4. 以往, 劇烈的氣候變化已襲擊地球多次。突發性的氣候變遷通常會持續好幾世紀甚至數千年。聯合國預測，往後的100年間，平均全球氣溫將升高1.5到4.5度C。

English Congruent Discourses

1. Pain has a strong psychological component.What the patient is thinking can determine his or her pain scale. Shifting patients' attention by, for example, having them listen to music can help reduce pain.
2. We already know that restricting an animal's calories can extend its life span. Even though this method was discovered 70 years ago, it is still the only

absolutely effective one. The restricted regime generally involves reducing an individual's food consumption by 30 to 40 percent.

3. About 85 to 90 percent of bad breath originates in the mouth. If healthy people have bad breath, the main source is their tongue, rather than their teeth. Apart from the mouth, another common source is the nose. In these cases, the odour has a very different quality.
4. In the past, dramatic climate changes have struck the earth many times. Sudden climate changes often persist for centuries or even thousands of years. The United Nations has predicted that, in the next 100 years, average global temperatures will rise 1.5 to 4.5 degrees C.

Chinese Incongruent Discourses

1. 很多人都有在家鄉迷路的經驗，因爲家鄉變化太大了。但是人口的成長卻隨著家庭成員減少而趨緩。因爲全球暖化，春天的花比50年前提早了一週綻開。
2. 一位高中生在一個科展的研究中創下連續十一天不睡的世界紀錄。好幾個其他的受試者在嚴密的監控下，也維持了8到10天清醒不睡的紀錄。
3. 許多人認爲污染是海洋物種減少的主因。有些人則很難相信漁獲不足，因爲他們當地的魚市場裡仍然可見成堆的魚。商業捕魚怎麼會對海洋物種有影響呢？
4. 在美國每年有三萬人自行結束生命。這個數字幾乎是去年愛滋病死亡人數的一半。在研究人員發展出有效的自殺預知方法之前，醫生最好是對自殺身亡者的親屬多加留意。

English Incongruent Discourses

1. Many people have the experience of getting lost in their hometowns because they have grown so much. But the population growth is slowing as families shrink. Because of global warming, spring flowers bloom a week earlier than they did 50 Years ago.

2. A high school student set his world record by being awake for 11 days in a science-fair project. Several other research subjects have remained awake for eight to ten days in carefully monitored experiments.
3. Many people thought that pollution is the main reason for the declines in marine species. Others find it hard to believe our shortage of fish because they still notice piles of fishes in their local fish markets. How could commercial fishing have any effect on the marine species?
4. Annually, 30,000 people in the U.S. take their own lives. That is roughly half of the number who died of AIDS last year. Until researchers can develop effective tests to forecast suicide, doctors may want to pay more attention to the relatives of suicide victims.

Appendix C

Material for Experiment 5-7

Incongruent Sentences

1. 老闆信任的工程師工作很認眞效率又高
2. 教授認識的作家很有名著作很多
3. 歌手讚美的指揮家很有天份但是很驕傲
4. 警衛拒絕的推銷員在門口招攬生意
5. 市長支持的議員整天講電話因爲事情很多
6. 女士遇見的女孩頭髮很長而且很漂亮
7. 伯伯拜訪的鄰居家境不好常需要幫忙
8. 偵探跟隨的警察很想知道眞相所以很積極
9. 小丑模仿的政客很受歡迎常被要求簽名
10. 編輯責備的科學家總是穿髒衣服
11. 助理追求的詩人充滿不切實際的幻想
12. 總統邀請的官員心懷不軌但是善於隱藏
13. 居民原諒的軍官受傷了被送往醫院
14. 司機抱怨的乘客總是很吵令人受不了
15. 主婦嘲笑的銀行家那天傍晚開跑車回家
16. 救火員發現的建築師在高級餐廳裡接電話
17. 飛行員訓練的技師脾氣不好對同事大吼大叫

18. 雇主喜歡的木匠拿著很多東西走路東倒西歪
19. 和尚聘請的義工看晚間新聞上的特別報導
20. 國王欣賞的將軍因爲醜聞自殺
21. 畫家鼓勵的詩人在他們的友情公開後寫了自傳
22. 學生質疑的老師在暑假寫了一本科學小說
23. 裁判批評的觀衆在比賽後公開談論意外
24. 校長面談的講師有一間很小的辦公室
25. 會計師連絡的市長開他的車去接小孩
26. 流氓攻擊的工人決定向他提出告訴
27. 演員尊敬的清潔工在蒐集獎券還有優待券
28. 犯人誤導的律師在打輸官司後換了工作
29. 音樂家嫁的漁夫上禮拜抓到一隻大鯊魚
30. 王子親吻的秘書付了派對的費用
31. 神父感謝的修女在雨中找到遮雨處
32. 農夫懷疑的專家在展覽會推銷產品
33. 侍者邀請的廚師嘗了肉的醬料
34. 護士協助的醫生爲手術借了器具
35. 導演開除的製作人提供問題的解決方法
36. 紳士挑戰的主席給了一場戲劇性的辭職演說
37. 球員迴避的教練進了健身房淋浴
38. 病人擁抱的牙醫掉了托盤在地上
39. 客戶幫助的藝術家開出一筆世紀交易
40. 小偷認得的間諜在紅綠燈過馬路
41. 員工糾正的經理修改了報紙的文章
42. 士兵槍擊的敵人因戰爭獲得一枚勳章
43. 富翁贊助的攝影師在旅程中弄壞了相機
44. 小孩欺騙的乞丐最後持有那筆錢
45. 總裁討厭的發言人在公司合併後賣了股份

46. 水手接納的船長擁有一家工廠跟一家餐廳
47. 顧客打擾的店員忘了櫃檯上的收據
48. 商人介紹的夥伴在會議裡發表報告

Congruent Sentences

1. 商人介紹的夥伴在會議裡發表報告
2. 老闆信任工程師他工作很認眞效率又高
3. 教授認識作家他有很名著作很多
4. 歌手讚美指揮家他很有天份但是很驕傲
5. 警衛拒絕推銷員他在門口招攬生意
6. 市長支持議員他整天講電話因爲事情很多
7. 女士遇見女孩她頭髮很長而且很漂亮
8. 伯伯拜訪鄰居他家境不好常需要幫忙
9. 偵探跟隨警察他很想知道眞相所以很積極
10. 小丑模仿政客他很受歡迎常被要求簽名
11. 編輯責備科學家他總是穿髒衣服
12. 助理追求詩人他充滿不切實際的幻想
13. 總統邀請官員他心懷不軌但是善於隱藏
14. 居民原諒軍官他受傷了被送往醫院
15. 司機抱怨乘客他總是很吵令人受不了
16. 主婦嘲笑銀行家他那天傍晚開跑車回家
17. 救火員發現建築師他在高級餐廳裡接電話
18. 飛行員訓練技師他脾氣不好對同事大吼大叫
19. 雇主喜歡木匠他拿著很多東西走路東倒西歪
20. 和尙聘請義工他看晚間新聞上的特別報導
21. 國王欣賞將軍他因爲醜聞自殺
22. 畫家鼓勵詩人他在他們的友情公開後寫了自傳
23. 學生質疑老師他在暑假寫了一本科學小說

24. 裁判批評觀衆他在比賽後公開談論意外
25. 校長面談講師他有一間很小的辦公室
26. 會計師連絡市長他開他的車去接小孩
27. 流氓攻擊工人他決定向他提出告訴
28. 演員尊敬清潔工他在蒐集獎券還有優待券
29. 犯人誤導律師他在打輸官司後換了工作
30. 音樂家嫁魚夫他上禮拜抓到一隻大鯊魚
31. 王子親吻秘書他付了派對的費用
32. 神父感謝修女他在雨中找到遮雨處
33. 農夫懷疑專家他在展覽會推銷產品
34. 侍者邀請廚師他嘗了肉的醬料
35. 護士協助醫生他爲手術借了器具
36. 導演開除製作人他提供問題的解決方法
37. 紳士挑戰主席他給了一場戲劇性的辭職演說
38. 球員迴避教練他進了健身房淋浴
39. 病人擁抱牙醫他掉了托盤在地上
40. 客戶幫助藝術家他開出一筆世紀交易
41. 小偷認得間諜他在紅綠燈過馬路
42. 員工糾正經理他修改了報紙的文章
43. 士兵槍擊敵人他因戰爭獲得一枚勳章
44. 富翁贊助攝影師他在旅程中弄壞了相機
45. 小孩欺騙乞丐他最後持有那筆錢
46. 總裁討厭發言人他在公司合併後賣了股份
47. 水手接納船長他擁有一家工廠跟一家餐廳
48. 顧客打擾店員她忘了櫃檯上的收據
49. 商人介紹夥伴她在會議裡發表報告

Appendix D

Material for Experiment 8

Ba version

1. 作家把困難克服了然後從新開始創作
2. 顧客把冰淇淋吃掉了然後點了另一個
3. 醫生把傷口割開了然後吸出有毒的血液
4. 校長把考試取消了因此他深受歡迎
5. 護士把藥片準備好接著她就去幫助病人了
6. 情報員把任務完成了然後他就去度假了
7. 導演把理想實現了然後他就退休了
8. 父母把小孩寵壞了結果老師感到很苦惱
9. 學生把收音機弄壞了結果老闆很生氣
10. 經理把會議忘掉了結果大家等了一個小時
11. 記者把健康忽略了結果生了一場大病
12. 市長把工廠拆除了然後興建一個博物館
13. 警察把罪犯逮捕了所以市民再也不用擔心
14. 銀行把利息降低了希望減少客戶存款意願
15. 總統把總理換掉了但是大多數議員反對
16. 專家把技術改良了因此得到大量訂單
17. 病毒把檔案刪除了甚至會破壞電腦

18. 工廠把成本降低了然後控制水與電
19. 垃圾把河流污染了因此居民開始清理河流
20. 教授把襯衫脫掉了因爲天氣實在太熱了
21. 大學生把時間浪費掉結果畢業後馬上後悔
22. 侍者把咖啡打翻了但是顧客一點也沒生氣
23. 遊客把森林燒掉了結果很多動物因此受傷
24. 病人把治療停止了但事實上復原機會很大

Non-Ba version

1. 作家克服了困難然後從新開始創作
2. 顧客吃掉了冰淇淋然後點了另一個
3. 醫生割開了傷口然後吸出有毒的血液
4. 校長取消了考試因此他深受歡迎
5. 醫生割開了傷口然後吸出有毒的血液
6. 情報員完成了任務然後他就去度假了
7. 情報員完成了任務然後他就去度假了
8. 父母寵壞了小孩結果老師感到很苦惱
9. 學生弄壞了收音機結果老闆很生氣
10. 經理忘掉了會議結果大家等了一個小時
11. 記者忽略了健康結果生了一場大病
12. 市長拆除了工廠然後興建一個博物館
13. 警察逮捕了罪犯所以市民再也不用擔心
14. 銀行降低了利息希望減少客戶存款意願
15. 總統換掉了總理但是大多數議員反對
16. 專家改良了技術因此得到大量訂單
17. 病毒刪除了檔案甚至會破壞電腦
18. 工廠降低了成本然後控制水與電
19. 垃圾污染了河流因此居民開始清理河流

20. 教授脫掉了襯衫因爲天氣實在太熱
21. 大學生浪費掉時間結果畢業後馬上後悔
22. 侍者打翻了咖啡但是顧客一點也沒生氣
23. 遊客燒掉了森林結果很多動物因此受傷
24. 病人停止了治療但事實上復原機會很大

Appendix E

Filler Sentences

Filler sentences

1. 院長於二月赴美訪問期間將進行國際合作
2. 這項物理學的重大發現影響人類的平均壽命
3. 諾貝爾獎得主成立獎學金鼓勵用功的學生
4. 年輕的學者已經指導了很多研究生
5. 學校買了很多植物準備種在花園裡
6. 所有的國會議員都支持總統的決定
7. 政府的投資明顯改善了地方建設
8. 公司開發了新技術因此賺了大錢
9. 由於預算有限公司必須減少支出避免浪費
10. 從下個月開始公園裡將要安裝風力發電機
11. 新的藥物據說可以治療頭痛
12. 有名的作家準備在記者會上發表新書
13. 學生特別喜歡在體育課看足球比賽
14. 外國人通常覺得使用筷子非常困難
15. 醫生建議我們每天要喝兩千毫升的水
16. 在汽油裡添加植物油對引擎不好
17. 工人不想放棄他的夢想繼續努力地存錢

18. 鄰居飼養一匹五個月大的騾子非常可愛
19. 每個人都會犯錯所以不需要太擔心
20. 電視台製作了大型海報吸引民衆注意力
21. 聯合國正在決定要如何處理人權危機
22. 鄰國應該要互相尊重不該侵犯對方領土
23. 日本與俄國首相簽署條約結束衝突
24. 強勁的颶風摧毀了重要設備工廠遭受重大損失
25. 旅行社突然更改計畫卻沒有通知旅客
26. 離開教室前應該要關燈節省能源
27. 耶誕節前很多人開始裝飾客廳還有花園
28. 病人家屬即將控告醫院要求賠償
29. 兩家公司都在蒐集對方的商品情報
30. 國家公園裡的道路因爲颱風嚴重損壞
31. 這家小公司已經從國外接到大量訂單
32. 羅馬帝國曾經征服過大不列顛
33. 植物可以吸收二氧化碳並釋放氧氣
34. 獵人現在的主要責任是保護野生動物
35. 最近原油價格再度升高破了紀錄
36. 總統在大選之前提出很多經濟發展政策
37. 技師更改了電腦設定機器馬上恢復正常
38. 牙醫詳細檢查牙齒找出牙痛的原因
39. 科學家正在研究如何準確測量人的智商
40. 音樂家在十年裡創作了超過三十首經典歌曲
41. 廚師在這道非常辣的菜裡混合了日本與泰國的配方
42. 美國拳擊手擊敗俄國拳擊手拿下世界冠軍
43. 考古學家花了數十年尋找神秘的地下神殿
44. 藝術家用大膽的顏色描寫自然景色
45. 低熱量飲食據說可以延長壽命

46. 高速鐵路節省了很多英國與歐陸的交通時間
47. 群衆正試圖阻止卡車進入雨林
48. 連鎖飯店決定賣掉資產結束營業

List of Figures