Bilingualism: Language and Cognition 6 (3), 2003, 201–211 © 2003 Cambridge University Press DOI: 10.1017/S1366728903001135

Basic skills in a complex task: A graphical model relating memory and lexical retrieval to simultaneous interpreting*

INGRID K. CHRISTOFFELS
ANNETTE M. B. DE GROOT
LOURENS J. WALDORP
University of Amsterdam

Simultaneous interpreting (SI) is a complex skill, where language comprehension and production take place at the same time in two different languages. In this study we identified some of the basic cognitive skills involved in SI, focusing on the roles of memory and lexical retrieval. We administered a reading span task in two languages and a verbal digit span task in the native language to assess memory capacity, and a picture naming and a word translation task to tap the retrieval time of lexical items in two languages, and we related performance on these four tasks to interpreting skill in untrained bilinguals. The results showed that word translation and picture naming latencies correlate with interpreting performance. Also digit span and reading span were associated with SI performance, only less strongly so. A graphical models analysis indicated that specifically word translation efficiency and working memory form independent subskills of SI performance in untrained bilinguals.

Simultaneous interpreting (SI) is a complex skill that involves speaking and listening in different languages at the same time. It requires comprehension of input, which involves the temporary storage and extraction of meaning of this input, while the interpreter is involved in formulating the meaning of an earlier part of the speaker's speech in the output language and producing yet another part (e.g., Gerver, 1976; Padilla, Bajo, Cañas and Padilla, 1995; De Groot, 2000; Rinne et al., 2000). Meanwhile, the interpreter is working under severe time constraints because the flow of input is continuous; the interpreter has no control over the input rate of the source speech. SI is, therefore, a very demanding cognitive skill. Not surprisingly, it is prone to error even under normal listening conditions and very sensitive to unfavorable conditions such as noise, faster than normal delivery of speech, poor pronunciation, and technical complexity of the speech (Gerver, 1976; Gile, 1997). Yet, high quality output is important if one considers that far reaching decisions may depend on the accuracy of interpreting in, for instance, international politics.

SI is probably one of the most complex language processing tasks imaginable (Frauenfelder and Schriefers, 1997). No other language task combines the need to comprehend and produce speech simultaneously, and to simultaneously command and control two languages. Understanding SI may be of great interest to cognitive psychologists and psycholinguists because of the demanding nature of SI in terms of online processing and control. Despite these intriguing characteristics of the task, experimental data on SI is scarce (see Christoffels and De Groot, in press, for a review). Since SI is an extremely complex cognitive skill, our approach in this study was to break SI down into its processing components to explore what basic cognitive subskills are important to SI. Of a larger set of possible subskills, we focused on the roles of basic language skills and of memory.

Language skills

In SI efficient language processing may be very important; when little attentional processing is required for certain aspects of language processing, more processing capacity will be available for other task components and the results of those processing steps are more quickly available for further processing. There is some evidence of a relationship between SI experience and efficient language processing. For example, Fabbro and Darò

* The authors like to thank Maya Misra, René Zeelenberg, Jeroen Raaijmakers and the reviewers of this manuscript for their helpful comments. We also thank Gerda Boven and Will Wintjes of the Faculteit Tolk-Vertaler, Hogeschool Maastricht, and Marisa Stoffers for rating SI performance. I. K. Christoffels and L. J. Waldorp were supported by grants (575-21-011 and 575-25-013) from the Netherlands Organization for Scientific Research (NWO) foundation for Behavioral and Educational Sciences. Portions of this research were presented at the 3rd International Symposium on Bilingualism in Bristol, UK, April 2001.

Address for correspondence
Ingrid K. Christoffels, Maastricht University, Faculty of Psychology, Department of Neurocognition, P.O. Box 616, 6200 MD Maastricht, The Netherlands
E-mail: *i.christoffels@psychology.unimaas.nl*

https://doi.org/10.1017/S1366728903001135 Published online by Cambridge University Press

(1995) reported greater resistance to the detrimental effects of delayed auditory feedback in students of SI than in monolingual participants. Fabbro and Darò concluded that these students of SI were more resistant to the interfering effects of delayed auditory feedback because they had developed high general verbal fluency and because they had learned to pay less attention to their own verbal output. Furthermore, Bajo, Padilla and Padilla (2000) compared the performance of interpreters, students of interpretation, bilinguals, and professionals from other fields on a categorization task. They showed that interpreters categorized non-typical exemplars of the categories faster than the students and the bilinguals. This finding indicates faster semantic access in interpreters than in other participant groups.

A specific aspect of language processing is the ability to quickly access and retrieve words. Although the word processing level may be considered too basic a level to be important in interpreting, which usually involves comprehension and production of complete discourse, there are reasons to believe that basic lexical retrieval efficiency may be relevant after all. For instance, if finding an appropriate word for a concept during SI takes a long time, it is likely that the interpreting process may break down due to the loss of valuable processing capacity and time. In other words, the time required to retrieve the word that corresponds to a given concept should be as short as possible. Even more important may be the time required to retrieve the translation of an input word. Although the interpreter usually does not attempt to literally translate each word from the source language into the target language, some literal word-to-word translation is likely to play a role in SI. In fact, although the details of the actual translation process during interpreting are unknown, several authors theoretically distinguish two presumably coexisting translation routes (or strategies) in SI: A transcoding route and a meaning-based route (e.g., Anderson, 1994; Fabbro and Gran, 1994; Isham and Lane, 1994; Paradis, 1994; De Groot, 1997, 2000; Massaro and Shlesinger, 1997; Bajo, Padilla and Padilla, 2000). The transcoding route involves literal transposition of words or multi-word units, whereas in meaning-based interpreting the meaning of the input is initially fully comprehended and conceptually represented. From this non-verbal conceptual representation of the input's meaning, production takes place in the target language (see Christoffels, De Groot & Kroll, 2003). Especially, if interpreting proceeds by means of transcoding, efficiency in the retrieval of translation equivalents may be crucial. Among other things, this study attempts to explore the relevance of these two specific aspects of basic language processing in SI: the retrieval of a word for a given concept and the retrieval of a translation equivalent for a given word.

Memory skills

Working memory is also likely to play a substantial role in SI. The working memory system is conceived of as a short-duration memory system that is capable of simultaneously storing and processing information. Working memory is known to play a significant role in normal language processing (e.g., Gathercole and Baddeley, 1993), but in SI we may expect it to be especially important. The delay between input and output in SI is on average about two to three seconds or four to five words (Treisman, 1965; Goldman-Eisler, 1972; Barik, 1973; Gerver, 1976; Anderson, 1994; Christoffels and De Groot, 2003). This factor alone is a reason to assume that interpreting is demanding in terms of storage capacity in memory. Moreover, in SI many processes have to take place simultaneously, placing heavy demands on the executive control function of working memory. It is generally assumed that working memory capacity is limited and that the processing and storage functions compete for this limited capacity (e.g., Just and Carpenter, 1992; Gathercole and Baddeley, 1993; Shah and Miyake, 1999). This trade-off between storage and processing may be a bottleneck in SI, where both functions are important (Gile, 1997).

Engle, Tuholski, Laughlin and Conway (1999) concluded from a factor analysis that short-term memory and working memory reflect separate but highly related constructs that are differentially related to higher order cognitive abilities. It seems sensible to at least distinguish between short-term memory and working memory in the sense that different tasks appear to primarily test either the storage capacity or the online processing resources of the memory system. For example, the digit span task assesses passive recall abilities and may be regarded as an indication of the participant's short-term memory capacity. Measures that tax both processing and recall may be regarded as measures of the participant's working memory capacity, where the focus is more on online processing. For example, the reading span task, introduced by Daneman and Carpenter (1980), combines temporary processing and storage demands. This type of task correlates more highly with measures of language comprehension and other language processing tasks than measures that tap only storage capacity (see Daneman and Merikle, 1996, for a meta-analysis).

Padilla et al. (1995) and Bajo et al. (2000) provided evidence that interpreting is indeed associated with efficient (working) memory skills. Padilla et al. compared experienced interpreters with two groups of student interpreters in the second and third year of their study and with non-interpreters. They administered a simple digit span task and a reading span task. According to both tasks, the interpreters had a higher average memory capacity than all other participant groups. In a similar experiment, Bajo et al. (2000) reported comparable

https://doi.org/10.1017/S1366728903001135 Published online by Cambridge University Press

results across interpreters, student interpreters, and non-interpreters.

A relevant question regarding interpreting is whether there are functional capacity differences in memory between the first and second language: when memory tasks are administered in a second language (L2) the results are often found to differ from performance in the first, native language (L1). For passive recall tasks, like the digit span task, part of the between-language difference in recall is due to differences in articulation rate between the L1 and L2. It seems, however, that short-term memory capacity may not be the same in the native and a second language, even when the differences in articulation rate are taken into account (Chincotta and Underwood, 1997, 1998).

A few studies have also investigated tasks that assess working memory, such as the reading span, in different languages. Harrington and Sawyer (1992) found differences in word span and digit span in two languages, but no difference in reading span. Service, Simola, Metsaenheimo and Maury (2002) reported data that suggest that working memory capacity may interact with language proficiency. L1 (Finnish) reading span was higher than L2 (English) reading span for a participant group that was relatively less proficient in the L2 (psychology students), but there were no differences for more advanced speakers of the L2 (students majoring in English). Service et al. concluded that sentence processing in a language that is not completely automated consumes extra working memory resources so that the reduction in overall capacity can be detected by working memory span tasks.

The present study

In the present study we focused on the role of two possible subskills of SI that are likely to be important: lexical retrieval and working memory. Unlike previous studies, we explored the relation of lexical retrieval and working memory with SI for participants without any prior experience in interpreting. In doing so, we capitalized on naturally occurring individual differences in interpreting 'talent'. If these basic skills are indeed important in SI, individual differences in lexical retrieval efficiency and working memory capacity should correlate with individual differences in SI performance.

To assess SI performance an SI task was administered in which the participants were instructed to translate an auditorily presented text into the target language online. In addition, two common psycholinguistic tasks were used that tap into language processing at the word level and give an indication of the efficiency of word retrieval: word translation and picture naming. In a word translation task, a word is presented in one language and the participant is asked to give its translation equivalent as quickly as possible in the target language. This task provides a measure of how quickly a word in the source language activates its counterpart in the target language and is subsequently produced. In a picture naming task a picture is presented for which the participant is to name the concept that it represents. This task gives an indication of the time needed to retrieve the word that names a concept and to produce it.

We assessed memory skills by administering a simple verbal digit span and a reading span task. As mentioned earlier, the digit span task is not as predictive for higher order language processing such as reading comprehension as, for example, the reading span task is (Daneman and Merikle, 1996). Similarly, Harrington and Sawyer (1992) obtained lower correlations between a simple word or digit span in L2 and L2 reading ability than between L2 reading span and L2 reading ability. In this study, we nevertheless included the digit span task to find out whether short-term memory capacity, as assessed by the digit span task, captures part of the memory requirements in interpreting. Finally, in contrast to Padilla et al. (1995), and Bajo et al. (2000), who only administered the reading span in L1, we administered two matched versions of the reading span task in two languages because working memory capacity in L2 may be more important in restraining processing in SI than capacity in L1.

Performance on the picture naming, translation, reading span, and possibly also the digit span tasks was expected to correlate with interpreting performance. We also performed a graphical models analysis. Such an analysis indicates by means of partial correlations what, if any, the interrelations between the variables are.

Method

Participants

Twenty-four students of the University of Amsterdam participated in return for either course credit or payment. All participants were native speakers of Dutch. Their mean age was 21 years. They had received at least six years of formal training in English as a second language, for about 3–4 hours a week, at secondary school. Subsequently, at university they had been required to read English since most textbooks are in English. The participants had also been exposed informally to English from early childhood, via film, television and music. We administered a language questionnaire designed to evaluate the language history of our participants. On a 7-point scale (1 = no knowledge of English and 7 = native level), the mean self-rating of their passive knowledge of English was 6.2; their mean rating of active knowledge was 5.7. None of the participants rated themselves lower than 5.

Tasks and stimuli

Lexical retrieval

The WORD TRANSLATION task was administered in both language directions separately, from English (L2) into

https://doi.org/10.1017/S1366728903001135 Published online by Cambridge University Press

Dutch (L1) and from Dutch into English. On each trial, a word was presented on the screen and the participant was asked to give its translation as quickly as possible. The stimuli were 72 English and 72 Dutch words. Two word characteristics were manipulated: word frequency and cognate status. Cognates are similar to their translation equivalents in the other language in both meaning and form (e.g., English *cat* – Dutch *kat*). The items from the four conditions (high frequency cognates, low frequency cognates, high frequency non-cognates, and low frequency non-cognates) were matched on both word length and word concreteness. Further information on these stimuli can be found in Appendix 1a. These four kinds of words were included to ensure the generalizability of our findings to different types of words.

The PICTURE NAMING task was administered in both English and Dutch. Pictures were presented on a screen and participants had to respond verbally as quickly as possible by giving the name of the depicted object. Seventy-two pictures were selected from the Snodgrass and Vanderwart (1980) set, which consists of black line drawings on a white background. As for the word translation task, both frequency and cognate status were manipulated. The name agreement percentages for the pictures of the four subsets of words were equally large, and word length was matched across subsets. Further information on the properties of the stimuli is given in Appendix 1b.

Both the word translation task and the picture naming task started with four practice stimuli that differed from the experimental stimuli. For each participant the stimuli were presented in a different random order. In each trial a fixation cross was presented for 500 ms, accompanied by a beep. A stimulus appeared 100 ms later and stayed on screen until the participant responded, but not longer than 7 s. The reaction time was measured by means of a sound-activated switch (voice-key). The experimenter typed the participant's response and triggered the next trial.

Memory

The READING SPAN task was administered in both English and Dutch. The participant read aloud sentences that were shown centered on the monitor while trying to remember the last word of each sentence. As soon as the participant read out the last word of a sentence, the experimenter triggered the next trial, a white screen appeared for 500 ms, a beep sounded, and then the next sentence appeared. The sentences were presented in increasing set sizes that consisted of two to five sentences. There were three series of each set size, which added up to 42 recall words per task. After presentation of the last sentence in each set, the participants were signaled on screen to write down all final words of the set they recalled. They were instructed not to start with writing down the last word presented to them. No other restrictions on recall order were given. The sentences of the English task were partly derived from Harrington and Sawyer (1992). The final words of the English and Dutch versions of the task were matched across the two languages on length (on average 4.1 letters for both English and Dutch) and frequency (on average ^{10}Log frequency = 3.38 for English and 3.37 for Dutch). Sentence length was between 11 and 13 words for both the Dutch and English version of the task (on average 11.8 words for English and 11.7 for Dutch). For each participant the total number of correctly recalled words was calculated.

The DIGIT SPAN task was administered in Dutch.[1] The participants were asked to repeat verbally presented sequences of an increasing number of (pseudo-)random digits (i.e., no immediate repetitions of the same digit and no obviously meaningful order) in the exact order of presentation and to start recalling them immediately after the last digit was read to them. The digits were read aloud at a rate of one per second, starting with a sequence of four digits. There were three series of each number of digits: after presentation of three series of four digits, the next sequence contained five digits. The task continued until the participant was no longer able to correctly recall one out of the three series. The largest number of digits recalled before an error made was taken as the participants' digit span.

Simultaneous interpreting

In the SIMULTANEOUS INTERPRETING task stimulus material was presented auditorily to the participant over headphones. The participants were asked to translate the English text as well as possible into Dutch and to start translating while listening. They were told that they should try to translate the meaning of what was being said rather than translate literally. Recordings that had been used as part of a medium level English listening high school exam in the Netherlands were adapted and used in two practice sessions of one and four minutes, respectively. The experimental SI task-recording was about 4.2 minutes long. It concerned a text on the science of face perception that we assessed to be easy to understand. After participation the participants reported no problems in understanding the text. The recorded text was spoken by a native speaker of British English at a rate of 116 words per minute on average. The interpreting output of the participants was recorded on the computer. From these recordings, two different measures of quality of SI performance were derived. The first measure was based on transcriptions of the recordings (SI-T). Ten sentences, approximately equally distributed in the text, were selected from the English text. Two judges rated the sentences on how well the content was translated into

[1] Practical constraints prevented presentation of the digit span in English.

https://doi.org/10.1017/S1366728903001135 Published online by Cambridge University Press

Dutch, on a scale from 0 to 3. A score of zero meant that the sentence was not translated at all; a score of 3 meant that almost all of the original text was translated into a proper grammatically correct Dutch sentence. For each participant a performance measure was calculated by adding the scores on the 10 sentences. The second measure (SI-A) was a rating of the translations' complete audio-recording by two judges, both teachers at the school for translation and interpreting of Maastricht, the Netherlands, and experienced simultaneous interpreters themselves. They rated each participant's performance on a scale from 0 to 5 on how well the text was interpreted. The inter-rater reliability ρ on both performance measures was reasonably high. For SI-T (transcription of the recordings) ρ was 0.91. For SI-A (recordings) ρ was 0.80.

Procedure and design

The participants were tested individually in a dimly illuminated room. All the instructions were in Dutch. They were provided on paper and administered verbally to the participants. After receiving a general instruction, a detailed instruction was provided for each separate task. The experiment was run on two different computers using three software packages. The naming and translation tasks were programmed to run on a Macintosh PPC 4400 with a 15-inch monitor, using the in-house developed software Flexi 4.0.5. All other tasks were run on a Macintosh G3 PowerBook. The memory span tasks were programmed using Psyscope 1.2 and the interpreting task was administered and recorded using DeckII^{TM}2.6.1, an audio-record program enabling us to simultaneously play the stimulus sound file while recording the participants' response on another track.

The session started with a brief voice-key practice in which participants read aloud English words. The complete session took about 1.5 hours. Apart from brief breaks between tasks, a break of five minutes was held halfway through the session. The tasks were administered in two different orders, each starting with a voice-key practice and ending with a short language questionnaire on the language background of the participants. Half of the participants performed the tasks according to the following task sequence: picture naming English; translation Dutch-English; reading span English; interpreting practice 1; interpreting practice 2; short break; digit span; reading span Dutch; translation English-Dutch; picture naming Dutch; interpreting task. The remaining half of the participants performed the tasks according to the same task sequence but they started with the Dutch version of the tasks instead of the English version.

Results and discussion

No main effect for order of presentation approached significance, so the data were collapsed across presentation orders. For the translation task and the picture naming task, the mean reaction time and percentage of errors for each of the conditions were calculated for each of the participants. All reaction times (RTs) longer than 5000 ms were treated as outliers and were excluded from the analysis. With this procedure less than 0.5% of the RTs was excluded from the data of both tasks. Furthermore, all RTs of incorrect responses and voice key errors were excluded. In all, less than 11% of the data in both tasks was excluded, of which about 7% were incorrect responses. Means and standard deviations on all tasks are presented in Table 1.

Table 1. *Means and standard deviations (SD) for all tasks.*

Tasks	Mean	SD
Word retrieval		
Translation English-Dutch	990	159
Translation Dutch-English	951	170
Picture naming English	1151	171
Picture naming Dutch	815	97
Memory		
Reading span English	32.9	3.73
Reading span Dutch	33.0	4.01
Digit span	6.4	1.21
SI performance		
SI-T (transcription)	18.2	3.91
SI-A (audio-recording)	2.3	0.86

Note. For translation and picture naming the reported averages are the reaction time for correct responses in milliseconds; for reading span, the number of correctly recalled words (maximum score = 42); for digit span, the number of correctly recalled series of digits; and for SI performance (SI-T and SI-A) the mean rating of the two raters (maximum score = 30 and 5, respectively).

The correlational and graphical models analyses, described next, are of main interest. However, for completeness, we briefly report the repeated-measures analyses of variance (ANOVA) that were conducted on the translation and picture naming latencies. Because the number of errors is relatively small, no separate analyses for the errors are presented. There was no speed–accuracy trade-off in either task, since the largest number of errors was found in the condition in which the participants reacted most slowly.

For the TRANSLATION task, within-subject factors were translation direction (English-Dutch, Dutch-English), cognate status (cognate, non-cognate), and frequency (high, low). The ANOVA yielded significant main effects for both cognate status ($F(1,23) = 118.97$, $p < .0001$) and frequency ($F(1,23) = 97.42$, $p < .0001$), which were qualified by an interaction between cognate status and frequency ($F(1,23) = 11.25$, $p < .005$). There was

https://doi.org/10.1017/S1366728903001135 Published online by Cambridge University Press

no significant main effect for translation direction ($F(1,23) = 2.81$, $p > .10$). Participants responded faster to cognates than to non-cognates, faster to high frequency words than to low frequency words and the latter difference was larger for the non-cognate words.

The repeated-measures ANOVA on the PICTURE NAMING task with the factors language, cognate status, and frequency yielded a significant main effect for language ($F(1,23) = 158.49$, $p < .0001$), cognate status ($F(1,23) = 48.40$, $p < .0001$), and frequency ($F(1,23) = 57.34$, $p < .0001$). Furthermore, there was an interaction between language and frequency ($F(1,23) = 30.52$, $p < .0001$). No other interactions approached significance. Participants were slower to name in English than in Dutch and the difference between high and low frequency words was larger when participants named in English than in Dutch.

Note that the effect of translation direction was not significant. It is often reported that forward translation, i.e., translation into the L2, is slower and more prone to errors than backward translation, into the L1 (e.g., Kroll and Stewart, 1994; Kroll, Michael, Tokowicz and Dufour, 2002). The absolute difference in RT between forward and backward translation actually suggests the opposite pattern: translation into English was, if anything, faster. Both null results and this opposite effect have been reported before (De Groot, Dannenburg and Van Hell, 1994; La Heij, Hooglander, Kerling and van der Velden, 1996; De Groot and Poot, 1997; Van Hell and De Groot, 1998; Christoffels, De Groot & Kroll, 2003). De Groot and Poot (1997) tested three proficiency groups. Two groups were secondary school pupils; the highest proficiency group consisted of university students. The former two groups were faster at forward translation; for the students the effect of translation direction was not significant. The equivalent RTs in both translation directions therefore suggests that our participants were relatively proficient in L2. Nevertheless, they were faster to name pictures in L1 (Dutch) than in L2 (English).

For the READING SPAN task, a paired t-test showed that there were no differences in performance between the Dutch and the English version ($t(1,23) < 1$, $p > .10$). In other words, functional working memory capacity of our participants appears to be equally large in L1 and L2, although the absolute difference was in favor of Dutch reading span. Harrington and Sawyer (1992) similarly reported no differences between languages on a reading span task. Furthermore, this pattern matches the advanced group in the Service et al. (2002) data, even though our participants were psychology students, like their less advanced group. This may, again, be taken to indicate that these participants were quite proficient in their L2.[2]

Table 2. *Correlation between the performance on all tasks and two measures of performance on the interpreting task, SI-T and SI-A.*

Task	SI-T	SI-A
Translation English-Dutch	$-.62^{***}$	$-.37^{*}$
Translation Dutch-English	$-.66^{***}$	$-.54^{**}$
Picture naming English	$-.62^{***}$	$-.57^{**}$
Picture naming Dutch	$-.41^{*}$	$-.44^{**}$
Reading span English	$.44^{*}$	.03
Reading span Dutch	.30∼	.24
Digit span	$.35^{*}$	$.38^{*}$

$^{*}p < .05$, $^{**}p < .01$, $^{***}p < .001$, $\sim 0.1 < p < .05$, one-tailed Z-test.

Note. Note that negative correlations are expected for the translation and picture naming tasks because better performance is associated with shorter reaction times and higher scores on the SI measures.

Correlational analyses

The correlations between performance on the language and memory tasks on the one hand, and the measures of simultaneous interpreting, SI-T and SI-A, on the other hand, are presented in Table 2. SI-T and SI-A correlated significantly with each other ($r = 0.86$, $p < .0001$), indicating that they were similar measures of SI performance. Nevertheless, the pattern of correlations between the two SI performance measures and the other tasks were somewhat different for the two measures.

On inspection, Table 2 shows that the translation task in each translation direction correlated quite highly with both interpreting measures. Furthermore, picture naming latencies correlated with interpreting performance although the correlations are somewhat stronger for picture naming in English. These results indicate that fast word retrieval is associated with better interpreting performance. In other words, some basic language skills may indeed be important for SI and fast word retrieval might be one of them.

[2] To support the idea that the lack of a difference between translation directions and between Dutch and English reading span was due to the relatively high L2 proficiency of our participants, we performed an additional analysis. The difference between forward and backward translation and the difference between English and Dutch reading span (i.e., the language effect) was calculated and correlated to the mean (passive and active) self-assessed proficiency in English. The correlation with the language effect in translation was significant ($r = -.51$, $p < .05$), indicating that a higher English proficiency was indeed associated with a smaller language effect. Similarly, the correlation of proficiency with the language effect in working memory was in the expected direction, albeit not significantly ($r = -.29$, $p > .10$). Higher proficiency is associated with a smaller difference between English and Dutch reading span, albeit not significantly so. Note that there was an extremely small range of proficiency scores (between 5 and 7 on a 7-point scale). The homogeneity of the participant group in this respect limited the power of this analysis.

https://doi.org/10.1017/S1366728903001135 Published online by Cambridge University Press

Reading span English correlated significantly with interpreting performance as assessed by SI-T, but not as assessed by SI-A. The correlation of reading span Dutch and SI-T showed a trend towards significance, but again, no significant relation is found with SI-A. For the digit span task the correlation with both interpreting performance measures was significant. These data give some indication that a larger or more efficient working memory is associated with better interpreting performance, and that also passive recall is associated with SI performance.

It thus seems that it is indeed possible to identify basic subskills that are relevant for SI performance. Better SI performance is associated with faster reaction times on word retrieval tasks. There is also some evidence that working memory is relevant, especially when considering SI-T. It is not clear why there are differences between the two measures of SI performance. The correlations were in general lower for SI-A. This might be because of the lower inter-rater reliability for SI-A ($\rho = 0.80$) than for SI-T ($\rho = 0.91$). Another difference between these measures is that they are based on different aspects of the output. SI-T is based on the transcription of the output, whereas SI-A is based on the recordings. The latter measure may be influenced more by aspects unrelated to how well the meaning of the input is preserved in the interpretations, such as interruptions in speech and type and clarity of voice. Further analysis is performed on SI performance according to SI-T because for this measure the inter-rater reliability was highest.

Graphical models analysis

In order to gain more understanding of the relation between the different variables in this study, we performed a graphical models analysis. This is a multivariate analysis that uses graphs to represent models and enables a representation of associations between variables (Edwards, 2000). The analysis is similar to path analysis. Graphical modeling is based on partial correlations between the variables. The procedure considers each of the individual edges (links between the variables) and subsequently removes those edges that are not required, i.e., for which the partial correlation does not differ significantly from zero. It is possible that variables become completely isolated, indicating that there is no association with any of the other variables. Graphical modeling is aimed at obtaining a model of associations that is as simple as possible and still describes the data well. With this analysis the relative importance of the associations between each of the tasks and SI can be assessed.

The analysis was performed with the MIM software package (Edwards, 2000). Using F-tests for stepwise backward testing on individual edges (starts from the full model with all edges present, by deleting edges), we obtained the model graphically presented in Figure 1. This model fits the data well, $\chi^2(21) = 14.117, p = 0.86$. Note that a model fits the data if the Chi-square test is not significant. The model was stable in the sense that two other model selection procedures yielded the same model: stepwise forward testing (starts from a model without any edges, by including necessary edges) and the Edwards-Havránek-search algorithm (Edwards, 2000). An analysis of the residuals indicated that the assumption of conditional multivariate normality was satisfied. It is important to emphasize that the model selection was in no way restricted by any a priori assumptions on the relations between the variables in this study.

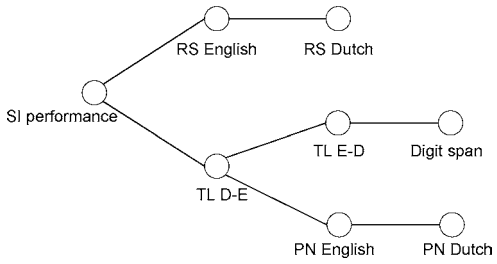


Figure 1. Graphical model of associations between SI performance and other tasks (RS = reading span, TL = translation task, PN = picture naming task).

From the graphical model we can make the following observations. First of all, there are only direct edges to SI performance from reading span English and translation Dutch-English. In other words, these two variables are the most relevant to SI because they have a direct effect on SI performance. Any influence of the remaining variables is mediated by reading span English or translation Dutch-English. Note that the absence of, for example, an edge between picture naming English or Dutch and SI performance means that picture naming in both languages is not directly associated with SI performance. In other words, although the ordinary correlation between picture naming and SI is significant (see Table 2), picture naming does not uniquely explain part of the variance in SI performance.

In the introduction, transcoding and meaning-based interpretation were discussed. The model emerging from the data indicates that transcoding may be a relatively important strategy for untrained SI. This is suggested by the fact that there is an edge between translation and SI, but not between picture naming and SI.

Second, a highly interesting result is that the model consists of two separate paths of tasks that are connected directly only to SI performance: on the one hand, the two working memory tasks are associated and, on the other hand, translation, picture naming, and digit span are associated. This means that we obtained evidence that different, (conditionally) independent subskills in SI exist,

https://doi.org/10.1017/S1366728903001135 Published online by Cambridge University Press

in the sense that working memory and translation each have an independent relation to SI performance.

Furthermore, in the model, the Dutch and English versions of each task are linked directly to each other. Thus, only one version of each task (i.e., reading span English, translation Dutch-English, picture naming English) is associated with other tasks, whereas the other language version (i.e., reading span Dutch, translation English-Dutch, picture naming Dutch) is connected in the model only through their counterpart. This was to be expected because the task may be assumed to capture the same underlying processes in the two languages. Therefore, one would not expect separate associations between, for example, both English-Dutch and Dutch-English translation with SI, because these tasks mainly explain the same variance. It is not completely clear why it is translation into English and not translation into Dutch that is directly linked to SI (which took place into Dutch). Although speculative, possibly the requirement to generate responses in L2 captures more of the restraints of processing in L2 that are relevant to SI than production in L1. However, one should take into account that the correlation between Dutch-English and English-Dutch translation is high ($r = .77$, $p < .05$), suggesting that, to a large extent, both tasks measure the same underlying skill. Moreover, Dutch-English translation is more easily selected in the model as the translation direction with a direct edge to SI, because this task's correlation with SI is slightly higher than the correlation of English-Dutch translation with SI (i.e., –.66 and –.62, see Table 1).

Finally, it is noteworthy that, in the model, digit span has no direct relation to SI performance. Apparently short-term memory capacity, at least in the native language, is not as relevant to SI as working memory is. Although this latter conclusion may not be surprising, perhaps it can be considered unexpected that digit span is associated with translation rather than with the working memory tasks. In this context it is worth mentioning that the correlations between digit span and word translation into Dutch ($r = –.59$) and into English ($r = –.50$) were higher than the correlation between digit span and reading span Dutch ($r = .41$) and English ($r = .22$), and that all correlations, except for the correlation with reading, span English reached significance. One suggestion concerning short-term memory may shed some light on why this pattern emerged. The digit span gives an indication of the capacity of the individual's phonological loop component. This slave-system of the working memory system is specialized for the retention of verbal material over short periods of time (e.g., Baddeley, 1986; Gathercole and Baddeley, 1993). Baddeley, Gathercole and Papagno (1998) argued that the primary function of the phonological loop is that of a language learning device (see also e.g., Ellis and Sinclair, 1996). It is found that vocabulary knowledge is strongly associated with digit span performance in children and with the learning of novel words in adults (e.g., Baddeley et al., 1998). Of the tasks in the present study, conceivably especially word translation skill is related to individual differences in foreign vocabulary acquisition ability. The relevance of the phonological loop in the acquisition of vocabulary may thus hint to why in our model the digit span task is associated with the translation task rather than the reading span task.

General discussion

Using correlational analysis and graphical modeling, we were able to show that it is possible to identify subskills in the complex task of SI. We found evidence that lexical retrieval and working memory are associated with SI performance.

The graphical model showed direct links from SI performance to reading span English and to translation into English. This indicates that of the subskills we studied, the efficiency of working memory and retrieval of translation equivalents are the most relevant subskills in SI. Interestingly, there are two (conditionally) independent paths to SI performance, namely working memory and word translation skills. This means that these subskills separately contribute to SI performance.

The relation between picture naming and SI was mediated by translation; there is no direct effect of picture naming on SI. It seems that, given knowledge about translation performance, individual differences in the efficiency of the processes underlying picture naming performance are not relevant to SI performance. Regarding memory skills, the English version of the reading span task had a direct effect on SI performance, which suggests that working memory capacity in L2 may be more important in constraining processing in SI than capacity in L1, even though average performance on the reading span did not differ for L1 and L2. Secondly, L1 digit span does not seem to capture any short-term memory requirements in SI on top of reading span. The implication of our model is that if we want to predict (L2 into L1) SI performance on the basis of the set of tasks included in this study, we would only need L2 working memory capacity and translation (into L2). The other tasks do not give any additional information on SI performance. This conclusion is important if we want to use these tasks in a practical setting, for example, to select or train SI students (note that we cannot be sure that our results generalize to L1-L2 interpreting since we only assessed interpretation into L1).

Given the complexity of the SI task, it is noteworthy that all our participants were able to produce at least some output when performing the task. It is likely that on a professional level SI requires particular skills or talents that not everyone possesses to the same degree.

https://doi.org/10.1017/S1366728903001135 Published online by Cambridge University Press

However, at a more modest level, SI seems to be within reach of reasonably proficient bilinguals untrained in SI, a fact that should be taken into account by current theories on (multilingual) language processing and memory.

Since the participants in this study were untrained in SI we cannot be certain that our findings generalize to professional interpreters. In fact, Neubert (1997) even claimed that untrained, or 'natural', translation is distinctly different from professional translation and interpreting. On the other hand, Harris and Sherwood (1978) stated that translation in general is an innate skill and that all bilinguals are able to translate within limits set by their proficiency (see also Malakoff and Hakuta, 1991; Malakoff, 1992). That interpreting skills may come about naturally is suggested by the fact that immigrant children often serve as interpreters for older members of their family (Harris and Sherwood, 1978). Although the literature does not provide a solid base on which to decide whether or not processing by professional interpreters is unique and differs from that of novices in this task, one study, reported by Dillinger (1994), suggests that at least comprehension processes in SI may be qualitatively the same for professionals and novices. The advantage of using, as we did, novices as participants is that this way the role of subskills is assessed without 'contamination' of the idiosyncratic strategies that interpreters may have developed (e.g., Shlesinger, 1994).

A final consideration is that our results might be explained by an underlying factor of general language proficiency in English. In other words, one might ask whether differences in general language proficiency generate the individual differences on both SI and the other tasks. The fact that in the model L1 tasks are not directly associated with SI but mediated by L2 may indicate that this is the case. Note, however, that language proficiency is a very general, multi-dimensional concept that involves pronunciation, grammar, vocabulary knowledge, and pragmatics, which can all be mastered at different levels by the L2 user. If one argues that word translation efficiency and English working memory capacity are different aspects of English language proficiency, we have at least still shown that these specific subskills are relevant.

However, there are in fact two findings that argue against the idea that global proficiency differences induce individual differences in performance on all tasks used in this study, including SI. First, the correlational analysis showed that L1 tasks also correlate significantly with SI performance, suggesting that the subskills are also relevant to SI if they are measured in L1. Moreover, in the graphical model two independent paths are associated with SI performance. That the working memory tasks were separated from the other tasks means that functional memory capacity in English and translation retrieval skills independently contributed to SI performance. This is a strong indication that not (only) individual differences in general English proficiency fuel the differences in performance on SI.

In future research it will be interesting to assess the role of different aspects of language proficiency and to include bilinguals of different proficiency levels to disentangle the relation between language proficiency and SI performance. It is also important to try to generalize our findings by including professional interpreters as participants. We have shown that efficient word translation and memory skills are important for SI by untrained bilinguals. Of course, other subskills may be relevant too. For example, interpreters have to speak only in the target language while they are listening to source language, which means that both languages have to be activated. It would, therefore, be interesting to assess whether individuals differ in the ability to control multiple languages without apparent interference from one language to another, and whether this ability is indeed related to SI. Also, in SI it would be useful if spoken language comprehension proceeded exceptionally fast; the ability to quickly comprehend discourse and resolve lexical and other types of input ambiguities may be another relevant subskill in SI (e.g., Gernsbacher and Shlesinger, 1997; Christoffels and De Groot, in press).

References

Anderson, L. (1994). Simultaneous interpretation: Contextual and translational aspects. In Lambert & Moser-Mercer (eds.), pp. 101–120.

Baayen, R. H., Piepenbrock, R. & van Rijn, H. (1993). *The CELEX Lexical Database* [CD-ROM]. Philadelphia: Linguistic Data Consortium; University of Pennsylvania [Distributor].

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

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

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

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

Chincotta, D. & Underwood, G. (1997). Bilingual memory span advantage for Arabic numerals over digit words. *British Journal of Psychology*, 88, 295–310.

Chincotta, D. & Underwood, G. (1998). Non-temporal determinants of bilingual memory capacity: The role of long-term representations and fluency. *Bilingualism: Language and Cognition*, 1, 117–130.

Christoffels, I. K. & De Groot, A. M. B. (2003). Components of simultaneous interpreting: Comparing interpreting with shadowing and paraphrasing. Ms., University of Amsterdam [submitted for publication].

https://doi.org/10.1017/S1366728903001135 Published online by Cambridge University Press

Christoffels, I. K. & De Groot, A. M. B. (in press). Simultaneous interpreting: A cognitive perspective. In J. F. Kroll & A. M. B. De Groot (eds.), *Handbook of Bilingualism: Psycholinguistic Approaches*. New York: Oxford University Press.

Christoffels, I. K., De Groot, A. M. B. & Kroll, J. F. (2003). Lexical retrieval and working memory in simultaneous interpreters: Expertise and language proficiency. Ms., University of Maastricht [in preparation].

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

Daneman, M. & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin and Review*, 3, 422–433.

Danks, J. H., Shreve, G. M., Fountain S. B. & McBeath M. K. (eds.) (1998). *Cognitive Processes in Translation and Interpreting.* Thousand Oaks: Sage Publications.

De Groot, A. M. B. (1997). The cognitive study of translation and interpretation. In Danks et al. (eds.), pp. 25–56.

De Groot, A. M. B. (2000). A complex-skill approach to translation and interpreting. In S. Tirkkonen-Condit & R. Jääskeläinen (eds.), *Tapping and Mapping the Processes of Translation and Interpreting*, pp. 53–68. Amsterdam: John Benjamins.

De Groot, A. M. B., Dannenburg, L. & Van Hell, J. G. (1994). Forward and backward translation by bilinguals. *Journal of Memory and Language*, 33, 600–629.

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

Dillinger, M. (1994). Comprehension during interpreting: What do interpreters know that bilinguals don't? In Lambert & Moser-Mercer (eds.), pp. 155–189.

Edwards, D. (2000). *Introduction to Graphical Modelling* (2nd edn.). New York: Springer-Verlag.

Ellis, N. C. & Sinclair, S. G. (1996). Working memory in the acquisition of vocabulary and syntax: Putting language in good order. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 49A, 234–250.

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

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

Fabbro, F. & Gran, L. (1994). Neurological and neuro-psychological aspects of polyglossia and simultaneous interpretation. In Lambert & Moser-Mercer (eds.), pp. 273–317.

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

Gathercole, S. E. & Baddeley, A. D. (1993). *Working Memory and Language.* Hove: Lawrence Erlbaum Associates.

Gernsbacher, M. A. & Shlesinger, M. (1997). The proposed role of suppression in simultaneous interpretation. *Interpreting*, 2, 119–140.

Gerver, D. (1976). Empirical studies of simultaneous interpretation: A review and a model. In R. W. Briskin (ed.), *Translation: Applications and Research*, pp. 165–207. New York: Gardner Press.

Gile, D. (1997). Conference interpreting as a cognitive management problem. In Danks et al. (eds.), pp. 196–214.

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

Harrington, M. & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition*, 14, 25–38.

Harris, B. & Sherwood, B. (1978). Translating as innate skill. In D. Gerver & H. W. Sinaiko (eds.), *Language Interpretation and Communication*, pp. 155–170. New York: Plenum Press.

Isham, W. P. & Lane, H. (1994). A common conceptual code in bilinguals: Evidence from simultaneous interpreting. *Sign Language Studies*, 85, 291–316.

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

Kroll, J. F., Michael, E., 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.

La Heij, W., Hooglander, A., Kerling, R. & van der Velden, E. (1996). Nonverbal context effects in forward and backward word translation: Evidence for concept mediation. *Journal of Memory and Language*, 35, 648–665.

Lambert, S. & Moser-Mercer, B. (eds.) (1994). *Bridging the Gap: Empirical Research in Simultaneous Interpretation*. Amsterdam: John Benjamins.

Malakoff, M. E. (1992). Translation ability: A natural bilingual and metalinguistic skill. In R. J. Harris (ed.), *Cognitive Processing in Bilinguals*, pp. 515–529. Amsterdam: North Holland.

Malakoff, M. & Hakuta, K. (1991). Translation skill and metalinguistic awareness in bilinguals. In E. Bialystok (ed.), *Language Processing and Language Awareness*, pp. 141–166. New York: Oxford University Press.

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

Neubert, A. (1997). Postulates for a theory of translation. In Danks et al. (eds.), pp. 1–24.

Padilla, P., Bajo, M. T., Cañas, J. J. & Padilla, F. (1995). Cognitive processes of memory in simultaneous interpretation. In J. Tommola (ed.), *Topics in Interpreting Research*, pp. 61–72. Turku: University of Turku.

Paradis, M. (1994). Toward a neurolinguistic theory of simultaneous translation: The framework. *International Journal of Psycholinguistics*, 10, 319–335.

Rinne, J. O., Tommola, J., Laine, M., Krause, B. J., Schmidt, D., Kaasinen, V., Teräs, M., Sipilä, H. & Sunnari, M. (2000). The translating brain: Cerebral activation patterns during

https://doi.org/10.1017/S1366728903001135 Published online by Cambridge University Press

simultaneous interpreting. *Neuroscience Letters*, 294, 85–88.

Service, E., Simola, M., Metsaenheimo, O. & Maury, S. (2002). Bilingual working memory span is affected by language skill. *European Journal of Cognitive Psychology*, 14, 383–407.

Shah, P. & Miyake, A. (1999). Models of working memory: An introduction. In A. Miyake & P. Shah (eds.), *Models of Working Memory: Mechanisms of Active Maintenance and Executive Control*, pp. 1–23. Cambridge: Cambridge University Press.

Shlesinger, M. (1994). Intonation in the production and perception of simultaneous interpretation. In Lambert & Moser-Mercer (eds.), pp. 225–235.

Snodgrass, J. G. & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 174–215.

Treisman, A. M. (1965). The effects of redundancy and familiarity on translating and repeating back a foreign and a native language. *British Journal of Psychology*, 56, 369–379.

Van Hell, J. G. & De Groot, A. M. B. (1998). Disentangling context availability and concreteness in lexical decision and word translation. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 51A, 41–63.

Received May 22, 2003
Revision received: July 25, 2003
Accepted July 25, 2003

Appendix 1a: ***Properties of stimuli used in the translation task***

	High frequency		Low frequency	
Measure	Cognates	Non-cognates	Cognates	Non-cognates
Dutch stimuli				
^{10}Log frequency	3.94	3.91	2.82	2.86
Cognate rating	5.46	1.39	5.39	1.39
Length (# of letters)	5.3	5.3	5.3	5.3
Concreteness	5.6	5.5	5.6	5.5
English stimuli				
^{10}Log frequency	3.89	3.91	2.89	2.87
Cognate rating	5.49	1.44	5.48	1.41
Length (# of letters)	5.3	5.3	5.3	5.3
Concreteness	5.5	5.5	5.5	5.5

Note. Dutch and English word frequency was taken from the CELEX-database (Baayen et al., 1993). The English frequency is multiplied by a factor of 2.26 to correct for the size difference between the Dutch and the English corpora. Cognate rating and concreteness were taken from De Groot et al. (1994). Analyses showed that statistically there were no differences between cognate score, frequency, word length, and word concreteness in the four word subsets other than the intended differences.

Appendix 1b: ***Properties of stimuli used in the picture naming task***

	High frequency		Low frequency	
	Cognates	Non-cognates	Cognates	Non-cognates
^{10}Log frequency	3.12	3.01	1.97	2.01
Cognate rating	5.8	1.6	5.8	1.7
Length (# of letters)	5.2	5.3	5.5	5.4
Agreement (%)	96	95	95	96

Note. English word frequency was taken from the CELEX-database (Baayen et al., 1993). Cognate ratings were obtained using a 7-point scale, where 1 represented least and 7 indicated most similarity in meaning and lexical form (n = 12). Agreement scores were taken from Snodgrass and Vanderwart (1980). Analyses showed that statistically there were no differences between cognate score, frequency, word length and agreement score in the four groups other than the intended differences.

https://doi.org/10.1017/S1366728903001135 Published online by Cambridge University Press