Available online at www.sciencedirect.com
SCIENCE DIRECT
Journal of Memory and Language 54 (2006) 324–345
Journal of Memory and Language
www.elsevier.com/locate/jml

Memory and language skills in simultaneous interpreters: The role of expertise and language proficiency ☆

Ingrid K. Christoffels [a,b,*], Annette M.B. de Groot [b], Judith F. Kroll [c]

[a] *Faculty of Psychology, Department of Cognitive Neuroscience, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands*
[b] *University of Amsterdam, The Netherlands*
[c] *Pennsylvania State University, USA*

Received 9 March 2005; revision received 30 November 2005
Available online 2 February 2006

Abstract

Simultaneous interpreting is a complex skill in which language comprehension and production take place at the same time in two languages. In this study, we examined performance on basic language and working memory tasks that have been hypothesized to engage cognitive skills important for simultaneous interpreting. The participants were native Dutch speakers proficient in English as a second language. We compared the performance of trained interpreters to bilingual university students (Experiment 1) and to highly proficient English teachers (Experiment 2). The interpreters outperformed the university students in their speed and accuracy of language performance and on their memory capacity estimated from a set of (working) memory measures. The interpreters also outperformed the English teachers, but only on the memory tasks, suggesting that performance on the language tasks was determined by proficiency more than cognitive resources. Taken together, these data point to (working) memory as a critical subskill for simultaneous interpreting.
Published by Elsevier Inc.

Keywords: Bilingualism; Translation; Simultaneous interpreting; Picture naming; Working memory; Expertise

☆ For assistance in collecting the data, the authors thank Katja Hoorn, Natascha Mooij, Els Oskam, Daphne Serpenti, Mirre Hubers, Paulien Ligthart, and Verena Schmittman. We thank Verena Schmittman, Lourens Waldorp, Jeroen Raaijmakers, Robert Hartsuiker and two anonymous reviewers for helpful comments on earlier drafts of this paper. This research was conducted while I. K. Christoffels was supported by Grant 575-21-011 from the Netherlands Organization for Scientific Research (NWO) foundation to A.M.B de Groot. The writing of this manuscript was supported in part by Grant (453-02-006) from the Netherlands Organization for Scientific Research (NWO) foundation to N.O. Schiller and in part by NSF Grants BCS-0111734 and BCS-0418071 to Judith F. Kroll at Pennsylvania State University. Portions of this research were presented at the 43rd Annual Meeting of the Psychonomic Society, Kansas City, Missouri, USA, and at the Third International Symposium on Bilingualism. Bristol, UK.
* Corresponding author. Fax: +31 43 3884125.
E-mail addresses: i.christoffels@psychology.unimaas.nl (I.K. Christoffels), a.m.b.groot@uva.nl (A.M.B. de Groot), jfk7@psu.edu (J.F. Kroll).

0749-596X/$ - see front matter. Published by Elsevier Inc.
doi:10.1016/j.jml.2005.12.004

Since the 1950s it has become common practice at large international meetings, such as those of the United Nations, to have interpreters simultaneously interpret into and from the different languages spoken by the members present. To an outside observer, simultaneous interpreting is a remarkable skill in that it places impressive demands on both language processing and memory. A question that immediately springs to mind is whether experienced interpreters possess special cognitive abilities that allow them to interpret successfully. Does acquiring skill in interpreting alter basic language processes in interpreters? Is their cognitive capacity to process language enhanced by virtue of their interpreting experience? Do they have unusual abilities to begin with? Although answers to these questions may provide critical information about the manner in which cognitive resources constrain language processing, there has been little experimental research on interpreting (see Christoffels, 2004; Christoffels & De Groot, 2005; for a review). The present study is one of the first steps towards understanding the way in which skill in simultaneous interpreting may be related to language processing and cognitive capacity.

From a cognitive perspective simultaneous interpreting is a striking task because of its sheer complexity. Many processes take place simultaneously. New speech input is presented continuously and the interpreter must comprehend this input and store it in memory. At the same time, earlier segments must be reformulated mentally into the target language and even earlier segments must be articulated (e.g., Gerver, 1976; Lambert, 1992; Padilla, Bajo, Cañas, & Padilla, 1995). The complexity of the task is illustrated by the fact that even professional interpreters sometimes make several mistakes per minute (Gile, 1997).

Perhaps the most salient characteristic of simultaneous interpreting is that interpreters must do simultaneously what ordinary language users typically do serially, that is, comprehend and produce language at the same time. In normal dialogue, some planning of speech during listening already takes place since people start their response a fraction of a second before their partner finishes speaking (Garrod & Pickering, 2004). However, during simultaneous interpreting output is continuously articulated. This simultaneity of comprehension and production is likely to be one important reason why interpreting is such a cognitively demanding task.

Another interesting characteristic of simultaneous interpreting concerns the time lag that necessarily occurs between the input and the corresponding output. Many language combinations differ in word order. Therefore, the interpreter may have to wait for the verb that may conclude a sentence in one language (e.g., German) but that needs to be produced early in the translation in another language (e.g., English). On the one hand, it is advantageous to wait as long as possible before starting to produce the translation of a given input utterance. The longer the lag, the more likely it is that potentially occurring ambiguities have been resolved by the context. On the other hand, when the lag is long there is more information to hold in memory than when the lag is short. Although the lag between input and output is influenced by factors such as the language of input (Goldman-Eisler, 1972), it is generally around 2 s or 4–5 words (Barik, 1973; Christoffels & De Groot, 2004; Gerver, 1976; Goldman-Eisler, 1972; Treisman, 1965).

A final critical aspect of simultaneous interpreting is that interpreters not only have to deal with the simultaneity of input and output, but also have to comprehend an utterance in one language but produce it in another. For participants without any previous experience in simultaneous interpreting the combination of simultaneity and language recoding was particularly demanding (Christoffels & De Groot, 2004). Thus, during simultaneous interpreting, both languages must be active simultaneously, although no language switches are allowed. This requires that interpreters ceaselessly control the use of the two languages concerned (e.g., Christoffels & De Groot, 2005; De Groot & Christoffels, in press).

The problem of control in bilingual processing becomes particularly critical when one considers the question of how bilinguals manage to keep their languages separated. Initial research on bilingual language representation suggested that word forms are represented separately for each language but that word meaning is shared between languages (Kroll & De Groot, 1997; Kroll & Stewart, 1994; Potter, So, Von Eckardt, & Feldman, 1984; Smith, 1997). More recently, studies have shown that information about word forms in both of the bilingual's languages is activated even when processing in one language. This recent evidence suggests that it is not possible to 'shut off' a language, even when it might be beneficial for task performance (e.g., De Groot, Delmaar, & Lupker, 2000; Dijkstra, Van Jaarsveld, & Ten Brinke, 1998; Jared & Kroll, 2001; Van Heuven, Dijkstra, & Grainger, 1998). The finding of nonselectivity has been shown to extend to production tasks, even when speakers intend to produce words in only one of their two languages (e.g., Colomé, 2001; Costa, Miozzo, & Caramazza, 1999; Hermans, Bongaerts, De Bot, & Schreuder, 1998). Evidence for cross-language syntactic priming indicates nonselectivity or integration between languages also on the syntactical level (Hartsuiker, Pickering, & Veltkamp, 2004; Loebell & Bock, 2003). If words from both languages are active during comprehension as well as production, then how is interference and switching into the nontarget language prevented?

One proposal to account for the control that bilinguals appear to exert in their speech is provided by the inhibitory control model (Green, 1986, 1998). According to this model, language selection requires inhibition of the candidates in the nontarget language. Achieving such control requires the allocation of attentional resources. Green (1998) argues that even a relatively simple task such as single word translation requires such control. Complex tasks such as online translation and interpretation may be assumed to require additional coordination of control processes (Gile, 1997).

We may ask whether experience in the complex bilingual task of interpreting that is hypothesized to impose high demands on cognitive control, enhances abilities that may be of particular importance to perform this task. A number of studies suggest that interpreters outperform other groups of participants on some language and memory tasks. Bajo, Padilla, and Padilla (2000) found that interpreters responded faster on atypical exemplars of categories in a semantic categorization task and that they were faster on nonwords in lexical decision, as compared to trainees learning to interpret, bilinguals without interpreting experience, and monolinguals. Furthermore, Fabbro and Darò (1995) observed that trainees learning to interpret were more resistant to the detrimental effects of delayed auditory feedback than participants with no simultaneous interpreting experience. Delayed feedback of a speaker's own voice typically causes speech disruptions. Padilla et al. (1995) also showed that interpreters demonstrated superior performance relative to other trainees and noninterpreters on digit and reading span tasks in the native language. Similar results were reported by Bajo et al. (2000). For example, unlike control groups, interpreters were not affected by articulatory suppression (i.e., continuous articulation of unrelated material) during recall (Padilla et al., 1995; Padilla, Bajo, & Macizo, 2005).

Although the advantages observed for interpreters in these studies could be due to the consequences of training in simultaneous interpreting or to self-selection factors that influence the choice to pursue interpreting as a career, the fact that they are observed even in the native language suggests that they may reflect more general characteristics of interpreters' cognitive abilities. There is also recent evidence, suggesting that a life time of bilingualism itself confers advantages to cognitive processing and executive control (e.g., Bialystok, 2005; Bialystok et al., 2005).

Research on participants without previous simultaneous interpreting experience also points to the relevance of language and memory skills to simultaneous interpreting. For these participants interpreting performance was correlated with measures of working memory and basic language skills (Christoffels, De Groot, & Waldorp, 2003) and to recall under articulatory suppression (Christoffels, in press).

Finally, off-line translation has been related to memory and language (Macizo & Bajo, in press). Manipulations of memory load, lexical ambiguity, and cognate status affected comprehension when reading for translation but not when reading for repetition (cognates are words that share aspects of both form and meaning across languages).

The goal of the present study is to begin to understand how basic components of language processing may be different when an individual is a skilled interpreter and how simultaneous interpreting is related to individual differences in memory capacity. In contrast to other studies, we assessed both native and second language capacity using different memory tasks. The approach we took was to compare the performance of different groups of bilinguals, all of whom were native speakers of Dutch (L1) and relatively proficient in English as a second language (L2), but who nevertheless differed in their proficiency in English and their professional training in simultaneous interpreting. If interpreting is a specific skill that does not affect the more basic components of language processing, then all of the participants should perform similarly on simple language processing tasks in their first and second languages. To increase sensitivity to possible qualitative processing differences, stimulus manipulations were included that might affect the groups differently. However, interpreters may differ from ordinary bilinguals in more than one way. Interpreters are necessarily highly proficient in L2 but, as suggested above, they appear to possess superior memory resources relative to other bilinguals. As noted, the latter may be a prerequisite for becoming an interpreter or a cognitive consequence of achieving skill in the simultaneous interpreting task.

In Experiment 1, we used a battery of individual difference tasks to assess the memory capacity of a group of simultaneous interpreters and a group of Dutch university students, all of whom were reasonably proficient in English. We then compared their performance on simple word translation and picture naming tasks. To anticipate the results, we show that the two groups differ on both the measures of working memory capacity and on their performance in the language processing tasks, with generally better performance for the interpreters than the bilingual university students. In Experiment 2, we tested a third group of participants who, like the interpreters, were highly proficient in English by virtue of being trained English teachers, but like the students, had no formal experience in interpreting. Here, the results show that language proficiency appears to be the main determinant of speed and accuracy on the language processing tasks; interpreters and highly skilled teachers showed similar performance on these tasks. However, on the working memory tasks, the teachers were no better than the university students, whereas the interpreters were superior to both groups. This finding suggests that there is a crucial relation between memory capacity and simultaneous interpreting skill.

Experiment 1: Interpreters and students

In Experiment 1 we focused on the role of working memory capacity and basic language processes. We compared experienced professional interpreters with Dutch university students who are relatively proficient in English as their L2. Note that this latter population has been studied extensively in past research and consists of unbalanced but proficient bilinguals (e.g., Dijkstra et al., 1998; Kroll & Stewart, 1994; La Heij, Hooglander, Kerling, & Van der Velden, 1996; Van Hell & De Groot, 1998b; Zeelenberg & Pecher, 2003) who are not as proficient as the interpreters. The most relevant difference between these two groups of participants was that they were either experienced in professional simultaneous interpreting or had never attempted to interpret before.

To assess whether simultaneous interpreting experience is associated with high cognitive capacity, we included several measures of working memory capacity. In addition to the word span task as a measure of passive recall, we included a reading span measure (Daneman & Carpenter, 1980). This is a concurrent memory task that is designed to tap both the demands of storage and the processing of information. According to Daneman and Green (1986), functional capacity varies with the processing characteristics of the task being performed. They proposed the speaking span task as a measure that taxes the processing and storage functions of working memory during sentence production. Since interpreting involves spoken language production, we also measured memory capacity with a speaking span task. Tests of memory capacity were administered in both languages because functional memory capacity may not be the same in the native and a second language (Chincotta & Underwood, 1998; Service, Simola, Metsaenheimo, & Maury, 2002).

In assessing basic language processing components, we specifically examined lexical retrieval. One might argue that lexical retrieval may be too basic a level to be relevant, since simultaneous interpreting usually involves the translation of larger units of discourse. However, it is likely that word retrieval skills are nevertheless relevant when performing the interpreting tasks successfully. Vocabulary problems may take up time and processing resources, which may eventually cause a breakdown in the simultaneous interpreting process. We presented two lexical retrieval tasks that have been used extensively in the past to study bilingual language representation: picture naming and word translation (e.g., Chen & Leung, 1989; De Groot, 1992; Kroll & Stewart, 1994; Potter et al., 1984). These tasks share a number of components, but differ in that in word translation lexical access precedes conceptual processing, whereas in picture naming this process is reversed (Miller & Kroll, 2002). In both the word translation and the picture naming tasks, we orthogonally manipulated two word type variables: word frequency and cognate status. These variables have produced robust effects in several studies that focused on the organization of bilingual memory (e.g., Costa, Caramazza, & Sebastian-Galles, 2000; De Groot & Nas, 1991; De Groot, 1992; Sanchez Casas, Davis, & Garcia Albea, 1992; Van Hell & De Groot, 1998a). We used these manipulations as a sensitive means to determine whether the nature of lexical processing differs across our participant groups. The effect of these manipulations may then also be replicated and extended to different types of participants. The effect of cognate status in picture naming is especially interesting because it indicates the influence of the language that is not the target in the task at hand.

To summarize, participants completed a battery of individual difference measures to measure memory capacity, including a word span task, a reading span task, and a speaking span task. Furthermore, they performed two on-line processing tasks, picture naming and single word translation. Both groups of participants performed all tasks in two languages, Dutch and English. Finally, we included two control tasks: a basic nonlinguistic reaction time test and an English vocabulary test. The reaction time test was included to control for differences between participant groups in basic reaction speed that are not related to language processing. The English vocabulary task was included to assess English proficiency in addition to self-ratings of the participants' proficiency level.

If the subskills examined here are indeed important to simultaneous interpreting, we predicted that the interpreters would outperform the students on both measures of language processing and memory capacity. On the control measures, we expected that the interpreters would have better vocabulary knowledge than the students, but that performance on the basic reaction time test should be unrelated to interpreting skill.

Method

Participants

One group of 40 participants consisted of university students who were unbalanced Dutch–English bilinguals. They spoke Dutch as their native language and were relatively proficient in English. They were undergraduate students of the University of Amsterdam and received course credit or a small monetary reward for participation. Of the 40 participants, 39 participants were included in the analysis, because the language questionnaire (see below) showed that one participant had some experience in (non-professional) interpreting. The participants had received at least six years of formal education in English as a second language, for about 3–4 h a

Table 1
Characteristics of participants in Experiments 1 and 2, including mean self-assessed proficiency in comprehension and production (rated between 1 and 10) in English

	Group		
	Students	Interpreters	Teachers
Rating of English production	7.3	8.3	8.1
Rating of English comprehension	7.9	8.7	8.6
Mean age (in yrs)	21.1	48.5	43.5
Professional experience (in yrs)	—	15.7	18.8

week, starting at age 12 at secondary school (often English lessons began at primary school). They use English textbooks in their university studies and have been exposed informally to English, via film, television, and music, in their daily lives.

The second group of participants consisted of 13 professional interpreters, all native speakers of Dutch. English was one of the languages they worked in professionally. They had on average 16 years of professional interpreting experience. All had attended university or other higher education, often specializing in language studies. Most of the interpreters had either attended a specialized interpreter course, or finished an internship, for example, at the European Union. We administered a language questionnaire designed to obtain more information about the language history of all of our participants. Participant characteristics and language questionnaire data are summarized in Table 1.

Materials and tasks

Word retrieval

For the picture naming task 72 pictures were selected from the Snodgrass and Vanderwart (1980) norms, which consist of black line drawings on a white background. Two word-type variables, word frequency, and cognate status, were manipulated orthogonally, forming four groups of words. Across the four subsets of words, word length was matched, and the percent name agreement (the percentage of participants using the same word to name a particular picture) was equally high and at least 90%.

The lexical characteristics of the stimuli are given in Tables 2 and 3. The picture naming task was administered separately in 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 entity. The task started with four practice pictures. On each trial, a picture was presented on the screen and the participant had to respond with its name as quickly as possible. The participants were instructed not to say anything else but the response. At the beginning of each trial a fixation cross was presented in the middle of the screen for 500 ms, accompanied by a beep. A picture 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 (that information was not visible to the participant) and triggered the next trial.

For word translation, 72 English and 72 Dutch words were selected. As in the picture naming task, frequency and cognate status were manipulated orthogonally. The four groups of words were matched on word length and word concreteness. The values on these variables were derived from a set of 440 words that had been rated on, among others, the variables relevant in this study (De Groot, Dannenburg, & Van Hell, 1994). The lexical characteristics of the words are given in Appendix 2. The word translation task was administered in both language directions separately, from English into Dutch and from Dutch into English. The task was administered in the same way as the picture naming task. It started with four practice words and reaction time was measured by means of a voice-key.

Memory tasks

All memory tasks were administered in both English and Dutch. The stimulus characteristics of the recall words are given in Table 4. For the reading span task two matched sets of 42 English and 42 Dutch sentences were developed. The sentences in the English test were partly derived from Harrington and Sawyer (1992). The final words of the sentences were matched on length and frequency across the two languages. Sentence length varied from 11 to 13 words for both the Dutch and English versions of the task (on average 11.8 words for English and 11.7 for Dutch). The sentences were randomly divided into three different series of sets of two, three, four, and five sentences, with the restriction that no rhyming final words were allowed within one set. The participant read aloud sentences that were shown centered on the monitor while trying to remember the last word of each sentence. The sentences were presented one by one on the computer screen in increasing set sizes. As soon as the participant finished reading the sentence, the experimenter triggered the next trial. An alert sounded and a white screen appeared for 500 ms, after which the next sentence appeared. After presentation of the last sentence in each set the participants were signaled by another alert beep and on screen to verbally recall the final words of each sentence in that set. The experimenter triggered the appearance of the next set of sentences when the participant finished recalling the current set. For each participant, the total number of correctly recalled words was calculated (maximum score: 42).

Table 2
Properties of stimuli used in the picture naming task

	High frequency		Low frequency	
	Cognates	Noncognates	Cognates	Noncognates
10 Log frequency	3.1	3.0	2.0	2.0
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, Piepenbrock, & van Rijn, 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.

Table 3
Properties of stimuli used in the translation task

Measure	High frequency		Low frequency	
	Cognates	Noncognates	Cognates	Noncognates
Dutch stimuli				
10 Log frequency	3.9	3.9	2.8	2.9
Cognate rating	5.5	1.4	5.4	1.4
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	3.9	3.9	2.9	2.9
Cognate rating	5.5	1.4	5.5	1.4
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 and concreteness rating 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.

Table 4
Properties of the words in the reading span, the speaking span, and the words span tasks

Task	10 Log frequency	Length (# of letters)
Reading span		
English	3.38	4.07
Dutch	3.37	4.07
Speaking span		
English	3.36	4.14
Dutch	3.37	4.10
Word span		
English	3.29	4.09
Dutch	3.37	4.22

Note. The English frequency as given in the corpus was multiplied by a factor of 2.26 to correct for the size difference between the Dutch and the English corpora. Analysis showed that statistically there were no differences in cognate score, frequency, word length, and word concreteness between the four word groups other than the intended differences.

For the speaking span task 42 words were selected for both the English and the Dutch version of this task. The English and Dutch words were matched on word frequency and length. The words were presented in three series of sets of two, three, four, and five words. Again no rhyming words were allowed within one and the same set. Participants were asked to (silently) read and remember the words. After presentation of a complete set the participant was asked to verbally produce a grammatically correct sentence for each of the words in the set that they recalled. The word appeared at the center of the screen for 1000 ms after a short alert. After 500 ms the next word was presented. The end of the set was signaled visually on the computer screen and by another sound. The experimenter triggered the appearance of the next set of words when the participant finished recalling and producing sentences for the current set. The total number of complete sentences containing the correctly recalled words was calculated (maximum score: 42).

For both the reading span and the speaking span task the participants were instructed not to recall the last words presented to them first. No other restrictions on recall order were given. Both tasks started with two sets of two practice items.

For the word span task 147 words per language were selected. The words were matched on word frequency and length across languages. The participants were asked to remember the words presented to them on screen and to recall them in exactly the same order. A 50 ms alert sounded, after which a word was presented for 1000 ms. Five hundred milliseconds after its offset the next word was presented. The words were presented in three series of four, five, six, up to ten words. The end of a series was signaled visually on screen and by another alert sound, at which point the participant started to recall the series of words. Presentation of the series of words was stopped when the subject failed to correctly recall one out of the three series of a given number of words. The number of correctly recalled series was calculated (maximum score: 21).

Control tasks

The vocabulary test was based on the English Vocabulary Test 10K (Meara, 1996). This test involves lexical decision for English words and nonwords without a speed instruction. The test does not include high frequency words and is meant to assess vocabulary knowledge at a relatively high level of proficiency (Cambridge proficiency level and above).

For the vocabulary test we used two subtests of the English Vocabulary Test 10K, which were administered digitally. Participants were asked to decide whether they knew the meaning of the words that were presented to them on screen and to indicate their decision by pressing one of two possible keys that indicated 'yes' or 'no' responses. The test consisted of 80 words and 40 nonwords. Each trial consisted of the presentation of a 500 ms fixation cross, a 250 ms empty screen, and a stimulus that appeared for a maximum of 6000 ms. The trial ended on response; the inter-trial time was 100 ms. The dependent variable was an index of number of 'hits' corrected for number of 'false alarms' (Huibregtse & Admiraals, 1999).

Finally, the arrow test was a basic reaction time test. In this task 40 leftwards and 40 rightwards pointing arrows were presented on the computer screen. Participants were asked to press a left hand key as quickly and as accurately as possible when an arrow was pointing to the left and a right hand key when an arrow was pointing to the right. First, a fixation cross was presented for 750, 1000, 1250, or 1500 ms, after which the arrow was presented (10 leftward and 10 rightwards pointing arrows for each duration). The arrow stayed on screen until a response was made but no longer than 5000 ms. The trials were presented in a different random order for each participant. The task started with 20 practice trials.

General procedure

All participants were tested individually in a dimly illuminated room, except for five interpreters who were tested individually in their own homes, because their schedules did not allow them to visit the laboratory. At the start of the session participants received both written and verbally presented general information on the plan for the experimental session. For each task a detailed instruction was provided. The experimental session began with voice-key practice. The complete session took about two hours. Apart from brief breaks between tasks, a five minute break occurred midway through the session. In and across tasks none of the stimulus words or picture names was repeated. The tasks were presented to the participants in two different orders, one that started with the Dutch version and another that started with the English version of the tasks: Voice-key practice; picture naming (English/Dutch), word span (English/Dutch), word translation (into English/into Dutch), reading span (English/Dutch), vocabulary task, speaking span (English/Dutch), short break plus interpreting task (not reported in this paper), word span (English/Dutch), basic reaction time test, speaking span (English/Dutch), word translation (into English/into Dutch), reading span (English/Dutch), picture naming (English/Dutch), and the language questionnaire.

Results

A multivariate analysis of variance showed that no effects of presentation order approached significance, so this factor was not taken into account in further analyses of the data. First, omnibus repeated-measures analyses of variance are presented on all tasks to establish the overall pattern of results. We then present specific analyses targeted at two questions. First, we asked whether there was a differential effect of language of the task on performance of each participant group, i.e., whether both students and interpreters reveal more skilled performance in their L1 than in their L2. Second, we examined group differences on each of the language versions of the tasks. We therefore present separate analyses for each language version.

All reported effects were tested against an α of 0.05. Because the sample of interpreters is not very large and, therefore, statistical power may be a concern, all p values under 0.10 were also reported. Except for the comparison between groups, which was the main focus on this study, effects are only reported when significant.

Word retrieval

Latency analyses are based on correct responses only. Data on trials in which reaction times were below 100 ms and above 5000 ms were considered outliers and removed from the analysis. This affected less than 0.5% of the data in all tasks in both languages (Dutch picture naming: 0.1%; English picture naming: 0.4%; Dutch–English translation: 0.4%; English–Dutch translation: 0.2%). Voice-key errors comprised 4.2% of the data for Dutch picture naming, 8.8% for English picture naming, 2.3% for Dutch–English translation, and 3.3% for English–Dutch translation.[1] In the overall analysis of variance, presented first, the main focus was to establish whether the participant groups differed in lexical retrieval latency. In these analyses the cognate and frequency (word type) manipulations are, therefore, not taken into account. In subsequent analyses by task, the word type manipulations are taken into consideration. Table 5 gives the F statistics and degrees of freedom associated with each source in the analyses. Planned contrasts and 95% confidence intervals using the mean-square error of the relevant interactions are used to assess effects of language and group and stimulus manipulations (Loftus & Masson, 1994; Masson & Loftus, 2003).

No detailed analyses of the errors are presented because in some conditions the percentage of errors is rather small. Main effects of group and language on the percentage of errors are reported to detect whether there are any speed-accuracy trade-offs.

Picture naming

The average picture naming reaction time on the correct responses and percentage of errors for the different groups of words are reported in Table 6 for both languages. The omnibus repeated-measures analysis of variance for the picture naming test with group (students vs. interpreters) as between-subjects factor and language (Dutch vs. English) as within-subjects factor showed that the groups differed marginally significantly from each other; the interpreters were faster than the students (881 and 975 ms, respectively, 95% confidence interval = 62 ms). Furthermore, a significant difference due

[1] On the basis of a stimulus analysis, additional analyses were performed on a subset of the stimuli. Stimuli were excluded when there were less than 50% correct responses across the students, the interpreters (Experiment 1), and the teachers (Experiment 2). This resulted in exclusion of one picture for Dutch picture naming, six pictures for English picture naming, one word for Dutch–English translation, and three words for English–Dutch translation (out of 72 stimuli per task). Removal of these items did not affect the outcome patterns of the statistical analysis.

Table 5
Analyses of variance results for Experiment 1: picture naming (Group × Language), picture naming Dutch (Group × Cognate status × Frequency), picture naming English (Group × Cognate status × Frequency), translation (Group × Language), translation English–Dutch (Group × Cognate status × Frequency), and Dutch–English (Group × Cognate status × Frequency)

Source	Degrees of freedom	F value	p
Picture naming			
Group	1, 50	3.48	.07
Language	1, 50	88.95***	<.001
Language × Group	1, 50	6.67*	.01
Picture naming Dutch			
Group	1, 50	.26	.62
Cognate status	1, 50	39.5***	<.001
Cognate status × Group	1, 50	.39	.54
Frequency	1, 50	50.90***	<.001
Frequency × Group	1, 50	1.44	.24
Cognate status × Frequency	1, 50	6.62*	.01
Cognate status × Frequency × Group	1, 50	2.03	.16
Picture naming English			
Group	1, 50	5.22*	.03
Cognate status	1, 50	63.05***	<.001
Cognate status × Group	1, 50	.02	.86
Frequency	1, 50	55.09***	<.001
Frequency × Group	1, 50	3.15	.08
Cognate status × Frequency	1, 50	3.59	.06
Cognate status × Frequency × Group	1, 50	3.18	.08
Translation (Group × Language)			
Group	1, 50	6.77*	.01
Language	1, 50	3.22	.08
Language × Group	1, 50	3.20	.08
Translation English–Dutch			
Group	1, 50	9.59**	<.01
Cognate status	1, 50	147.74***	<.001
Cognate status × Group	1, 50	.01	.92
Frequency	1, 50	94.55***	<.001
Frequency × Group	1, 50	.96	.33
Cognate status × Frequency	1, 50	25.87***	<.001
Cognate status × Frequency × Group	1, 50	.90	.35
Translation Dutch–English			
Group	1, 50	3.22	.08
Cognate status	1, 50	126.04***	<.001
Cognate status × Group	1, 50	.39	.54
Frequency	1, 50	95.75***	<.001
Frequency × Group	1, 50	.62	.44
Cognate status × Frequency	1, 50	24.05***	<.001
Cognate status × Frequency × Group	1, 50	5.56*	.02

* $p < .05$.
** $p < .01$.
*** $p < .001$.

Table 6
Picture naming: Mean reaction time (RT) in ms and percent errors (% error) for high frequency (HF) and low frequency (LF) cognate and noncognate words per language and per group, and the difference in reaction time between the two languages (language effect) for the students (Experiment 1), the interpreters (Experiment 1 and 2), and the teachers (Experiment 2)

Group	Language				Language effect
	Dutch		English		
	RT	% error	RT	% error	RT
Students (n = 39)					
Cognates					
HF	707	1.0	925	2.3	
LF	808	3.8	1183	9.8	
Noncognates					
HF	800	4.0	1117	15.5	
LF	879	5.4	1381	39.0	
Average	799	3.6	1152	16.7	−353
Interpreters (n = 13)					
Cognates					
HF	679	0.4	842	0.9	
LF	782	3.4	917	4.7	
Noncognates					
HF	818	5.1	961	8.5	
LF	843	8.1	1206	11.1	
Average	781	4.3	982	6.3	−201
Teachers (n = 15)					
Cognates					
HF	710	1.1	762	1.9	
LF	812	4.4	951	5.6	
Noncognates					
HF	789	4.4	943	4.1	
LF	966	3.0	1181	11.5	
Average	819	3.2	959	5.8	−140

to language with faster reaction times for Dutch than for English (790–1067 ms, 95% confidence interval = 36) was qualified by an interaction between language and group. Comparison against the 95% confidence interval for the students (42 ms) and the interpreters (72 ms) showed that not only the students but also the interpreters were faster in Dutch than in English picture naming. The disadvantage of naming pictures in the L2 was, however, larger for the students ($\Delta = 353$ ms) than for the interpreters ($\Delta = 201$ ms).

Separate analyses of variance were conducted for the Dutch and the English version of the task, with group as a between factor and cognate status (cognates vs. noncognates) and frequency (high frequency words vs. low frequency words) as within factors. For Dutch picture naming there was no statistically significant difference in latencies between groups, 781–799 ms (95% confidence interval = 61 ms). There were significant differences due to cognate status (744 ms for cognates and 835 ms for noncognates, 95% confidence interval = 25 ms), frequency (751 and 828 ms for high and low frequency, 95% confidence interval = 18 ms), and a significant interaction between cognate status and frequency. The cognate effect is larger for high frequency words ($\Delta = 103$ ms) than for low frequency words ($\Delta = 51$ ms). The 95% confidence interval for each of these differences was 21 ms. Noteworthy is that we obtained no significant interactions of these factors with group, indicating that there were no differences in performance between the students and the interpreters in Dutch picture naming.

For English picture naming the pattern of results was quite different. In English we did find a significant difference in latencies between groups. The interpreters were faster than the students in English picture naming (981–1152, 95% confidence interval = 129 ms).

The same effects of the word manipulations emerged in English as in Dutch: significant faster reaction times due to cognate status (967 ms for cognates and 1166 ms for noncognates, 95% confidence interval = 44 ms), frequency (961 and 1172 ms for high and low frequent words, 95% confidence interval = 49 ms) and a marginally significant interaction. The cognate effect was smaller for high frequency words ($\Delta = 156$ ms) than low frequency words ($\Delta = 189$ ms). The 95% confidence interval for each of these differences was 45 ms.

The error analysis showed that the percentage of errors was higher for the students than for the interpreters, 10.1–5.3% (95% confidence interval = 1.6%, $F(1, 38) = 11.18$, $p = .002$), that more errors were made in the L2 than the L1, 11.5–3.9% (95% confidence interval = 1.6%, $F(1, 50) = 34.16$, $p < .001$), and that this was especially the case for the students (students: $\Delta = 13.1\%$, 95% confidence interval = 1.8%, interpreters: $\Delta = 2.0\%$, 95% confidence interval = 3.2%, $F(1, 50) = 34.16$, $p < .001$). This pattern of results corresponds to the latency data.

Word translation

The mean correct translation reaction times and percentage of errors for each group of words in the two translation directions are reported in Table 7. The omnibus repeated-measures analysis of variance, with group (students vs. interpreters) as a between-subjects factor and language (Dutch–English translation vs. English–Dutch translation) as a within-subjects factor, showed that the latencies between groups differed significantly from each other. Reaction times were faster for interpreters than for students, 817–945 ms (95% confidence interval = 61 ms). Furthermore, both the difference due to translation direction, and the interaction between language and group were marginally significant. Planned

Table 7
Word translation: Mean reaction time in ms (RT) and percent errors (% error) for high frequency (HF) and low frequency (LF) cognate and noncognate words per translation direction, and the difference between the two languages in reaction time (language effect) for the students (Experiment 1), the interpreters (Experiments 1 and 2), and the teachers (Experiment 2)

Group	Language				Language effect
	English–Dutch		Dutch–English		
	RT	% error	RT	% error	RT
Students (n = 39)					
Cognates					
HF	784	1.3	712	1.1	
LF	944	8.7	904	3.7	
Noncognates					
HF	959	5.0	883	3.8	
LF	1225	29.1	1149	29.9	
Average	978	11.0	912	9.6	66
Interpreters (n = 13)					
Cognates					
HF	656	1.3	656	.43	
LF	753	0.9	745	.43	
Noncognates					
HF	803	2.1	782	1.3	
LF	1055	5.1	1083	12.8	
Average	817	2.35	817	3.74	0
Teachers (n = 15)					
Cognates					
HF	657	0.4	663	0.0	
LF	738	1.5	721	1.1	
Noncognates					
HF	763	0.7	769	0.0	
LF	992	4.1	986	12.6	
Average	787	1.7	785	3.4	2

contrasts revealed only for the students an effect of language. The students translated faster from Dutch into English, than from English into Dutch[2] ($\Delta = 66$ ms). For the interpreters, translation was equally fast in both directions of translation ($\Delta = 0$ ms). The 95% confidence intervals for the difference between means were 45 ms for the interpreters and 26 ms for the students.

Separate repeated-measures analyses of variance were conducted for both translation directions with group as a between subjects-factor and cognate status (cognates vs. noncognates) and word frequency (high frequency vs. low frequency) as within-subjects factors. These analyses revealed that for English–Dutch translation, the interpreters were faster than the students, 817–978 ms (95% confidence interval = 91 ms). Also for Dutch–English translation the interpreters were faster than the students, 817–912 ms (95% confidence interval = 93 ms), although for Dutch–English translation this difference was marginally significant and qualified by a significant three-way interaction of group, cognate status, and frequency. Both in English–Dutch translation and in Dutch–English translation there were significant differences due to cognate status (English–Dutch: 784 ms for cognates compared to 1010 ms for noncognates, 95% confidence interval = 32 ms, and Dutch–English: 755 ms for cognates compared to 974 ms for noncognates, 95% confidence interval = 35 ms). There was also a significant difference due to frequency (English–Dutch: 800 and 994 ms for high and low frequency, 95% confidence interval = 35 ms; Dutch–English: 758 and 970 ms for high and low frequency, 95% confidence interval = 38 ms) and a significant interaction between cognates status and frequency. For English–Dutch translation the effect of cognate status was smaller for high frequency words ($\Delta = 161$) than for low frequency words ($\Delta = 291$). The 95% confidence interval for each of these differences was 31 ms. For Dutch–English translation the 95% confidence interval for these differences was calculated separately for the interpreters (66 ms) and for the students (38 ms). For both groups the effect of cognate status was smaller for high—than for low frequency words (interpreters: $\Delta = 136$ and $\Delta = 337$ ms, respectively; students: $\Delta = 171$ and $\Delta = 245$ ms, respectively).

The percentage of errors was significantly higher for the students than for the interpreters, 10.5–3.0% (95% confidence interval = 1.9%, $F(1,50) = 21.26$, $p < .001$), indicating that the group differences were not due to a speed-accuracy trade-off. There was no difference in errors between translation directions, $F(1,50) < 1$, $p > .10$, but this factor interacted with group, $F(1,50) = 21.26$, $p = .047$. The students made slightly fewer errors ($\Delta = 1.35\%$, 95% confidence interval = 1.84) and the interpreters made slightly more errors ($\Delta = 1.4\%$, 95% confidence interval = 3.18) in Dutch–English translation than in the reverse translation direction.

Memory

Repeated-measures analyses of variance were conducted for the memory tests with group (students vs. interpreters) as between-subjects factor and language (Dutch vs. English) as within-subjects factor. Table 8 gives the F statistics and degrees of freedom associated with each source in the analyses. Planned contrast and

[2] Note that this effect of translation direction was opposite to what is predicted by the revised hierarchical model (Kroll & Stewart, 1994). Repeatedly an advantage is observed for translation into the L1 (e.g., Kroll, Michael, Tokowicz, & Dufour, 2002). However, null-effects, or faster translation into the L2 are not isolated findings (Christoffels et al., 2003; De Groot et al., 1994; De Groot & Poot, 1997; La Heij et al., 1996; Van Hell & De Groot, 1998b).

Table 8
Analyses of variance results for Experiment 1: reading span (Group × Language), speaking span (Group × Language), and word span (Group × Language)

Source	Degrees of freedom	F value	p
Reading span			
Group	1, 50	4.76*	.03
Language	1, 50	7.15*	.01
Language × Group	1, 50	2.12	.15
Speaking span			
Group	1, 50	16.23***	<.001
Language	1, 50	19.78***	<.001
Language × Group	1, 50	6.46*	.01
Word span			
Group	1, 50	14.66***	<.001
Language	1, 50	.36	.55
Language × Group	1, 50	5.24*	.03

* $p < .05$.
*** $p < .001$.

95% confidence intervals (Loftus & Masson, 1994; Masson & Loftus, 2003) are used to assess the difference between the Dutch version and the English version of the tasks (effect of language) for each participant group separately, followed by a contrast designed to assess the effect of group when tests were administered in the L1 and in the L2. The average performance of the interpreters and the students on the memory tasks is presented in Table 9.

Table 9
Reading span, speaking span, and word span

Task	Language		Language effect
	Dutch (L1)	English (L2)	
Reading span			
Students	34.00	31.13	2.87
Interpreters	35.39	34.54	.85
Teachers	32.06	30.73	1.33
Speaking span			
Students	29.00	24.77	4.23
Interpreters	32.08	30.92	1.16
Teachers	28.80	25.60	3.2
Word span			
Students	3.59	3.05	0.54
Interpreters	5.00	5.92	−0.92
Teachers	3.80	2.40	1.4

Mean number of correctly recalled words on the reading span and the speaking span tasks (maximum score = 42), and mean number of correctly recalled series of words in the word span task per language for students (Experiment 1), interpreters (Experiments 1 and 2), and teachers (Experiment 2).

Reading span

For the reading span task, a repeated-measures analysis of variance revealed statistically significant differences between groups, indicating that performance was better for the interpreters than for the students, 34.96–32.56 (95% confidence interval = 1.35). Furthermore, a significant difference due to language revealed better performance for the Dutch (34.69) than the English (32.83) version of the task (95% confidence interval = .86). The interaction between group and language was not significant. The 95% confidence interval for the difference between the means for English and Dutch was 1.71 for the interpreters and .99 for the students. For the interpreters there was no significant effect of language ($\Delta = .85$). In contrast, the students performed better on the Dutch version than on the English version ($\Delta = 2.87$). Furthermore, the 95% confidence interval for the difference between the means of the groups (1.35) suggests that the interpreters were significantly better than the students on both the English version, ($\Delta = 3.41$), and the Dutch version ($\Delta = 1.39$).

Speaking span

For the speaking span task we obtained significant differences in latencies between groups (interpreters: 31.50 and students: 26.89, 95% confidence interval = 1.41) and languages (Dutch: 30.54, and English: 27.85, 95% confidence interval = .76). Moreover, there was an interaction between group and language. The 95% confidence interval for the difference between the means for the English and Dutch version for the interpreters (1.49) and for the students (.86) showed that for the difference between the Dutch version and the English version of the task was not significant for the interpreters ($\Delta = 1.16$), whereas the students performed better on the Dutch version than on the English version ($\Delta = 4.23$). The difference between groups persisted in each language separately, the interpreters performing better than the students on both the English version ($\Delta = 6.15$), and the Dutch version ($\Delta = 3.08$). The 95% confidence interval for the difference between the means of the groups was 1.41.

Word span

For the word span task we obtained a significant difference in performance between groups (interpreters: 5.46, and students: 3.32, 95% confidence interval = .69), and an interaction between group and language. The 95% confidence interval for the difference between the means for the English and Dutch version for the interpreters (.78) and for the students (.45) showed that, unexpectedly, the interpreters performed worse on the Dutch version than on the English version of the task ($\Delta = .92$). The students performed better on the Dutch version ($\Delta = .54$). The 95% confidence interval for the difference between the means of the groups is .69. On both the Dutch and the English versions, the interpreters performed better than the students ($\Delta = 1.41$ ms and $\Delta = 2.87$ ms, respectively).

Control tasks

Average scores on the vocabulary test were .60 for the students and .86 for the interpreters. On the arrow test, the average reaction times were 347 ms for the students and 418 ms for the interpreters. Independent samples *t* tests were conducted on performance for each of these tasks. The analyses showed that interpreters performed better on the vocabulary test, $t(50) = 8.45$, $p < .001$, but that the students outperformed the interpreters on the arrow reaction time test, $t(50) = 6.39$, $p < .001$. The arrow test is therefore the only test in our experiment on which the students outperformed the interpreters. This difference may be due an effect of age on speed of processing, attributable to the fact that the interpreters were approximately 20 years older than the university students (see Table 1).

Summary and discussion

The main results for the word retrieval tasks can be summarized as follows. On both the translation and picture naming tasks the interpreters were faster than the students. In picture naming this group effect was present only in English. The only linguistic task in which the students did not differ significantly from the interpreters was Dutch (L1) picture naming. Fast lexical retrieval seems, therefore, to be a relevant subskill of simultaneous interpreting

Given that the interpreters were slower on the basic reaction time task it is especially noteworthy that they were faster overall in word retrieval. The relative slowness of the interpreters in the arrow task raises the possibility that if interpreters were to be matched in terms of age with a bilingual group like the students, also on this task the interpreters might actually be faster. Another possibility is that picture naming in the L1 is already at ceiling in all participants so that interpreting experience does not influence retrieval times in the L1. Note that the interpreters are as fast as the students on Dutch picture naming and faster on the other lexical retrieval tasks.

As expected, picture naming in the L2 was in general slower than picture naming in the L1. Perhaps surprising, this is also the case for the professional interpreters, who are presumably highly proficient in their L2. The difference between L1 and L2 picture naming was larger for the students. These results show that in the picture naming task an effect of language dominance occurs even in highly proficient bilinguals (the results also clearly indicate that the interpreters were dominant in Dutch). At the same time there was no effect of translation direction in the word translation task in this group. For the students on the other hand, it did matter what direction they translated in. They translated faster from Dutch into English than from English into Dutch.

Comparisons between students and interpreters on the (working) memory span tasks showed a clear pattern: The interpreters outperformed the students on all three memory tasks. Interestingly, the advantage of the interpreters over the students also held in Dutch, although this difference was not significant for the reading span task. Finally, the students' performance was affected negatively by being tested in their L2, whereas for the interpreters the language they were tested in had no effect on their performance. Memory skills thus appear to be relevant for simultaneous interpreting.

In all, this study showed clear differences between the students and the interpreters on all of the measured subskills that we assumed to underlie successful interpreting. However, the two groups we tested differ in more than one way from each other. Most importantly, they do not only differ in the amount of interpreting experience they have, they also differ in their L2 language proficiency. Therefore, the differences between the students and the interpreters may be due to general differences in English (L2) proficiency, rather than to differences in interpreting experience. However, the fact that, at least for the memory tests, we found an advantage of the interpreters over the students when they were tested in their L1 speaks against such an interpretation of our findings. It suggests that the differences between groups cannot be explained by language proficiency alone.

The exact role of language proficiency in performance in our tasks remains unclear. This is because it is difficult to tease apart the roles of language processes and capacity. The reason is that, on the one hand, the capacity measures are language based as well, whereas on the other hand, there is ample evidence that cognitive resources affect language processing (e.g., Daneman & Green, 1986; Daneman & Merikle, 1996; Gathercole & Baddeley, 1993; Hartsuiker & Barkhuysen, 2006; Just & Carpenter, 1992) and may be particularly critical in processing the L2 (e.g., Kroll et al., 2002; Michael & Gollan, 2005; Miyake, 1998).

In the next experiment, we again assessed bilingual performance on lexical retrieval and memory tasks, using a new group of bilinguals who were more proficient in their L2 than the bilingual university students, but who, as the university students, had no previous experience or training in simultaneous interpretation.

Experiment 2: Interpreters and Dutch teachers of English

In this experiment, we tested native Dutch speakers who were teachers of English. We used the same battery of tasks on which interpreters and bilinguals were compared in Experiment 1. We chose to include teachers of English because they are similar to the professional interpreters in several important ways. Most critically, they are likely to have obtained an equally high level of proficiency in English as a second language. Furthermore, they are more similar in age to the interpreters and share a more similar educational background—participants in both groups often held a degree in English. They share an interest in language in general, and in English in particular, and are professionally involved in English on a daily basis. The crucial difference between the two groups is the amount of simultaneous interpreting experience they have had: none for the teachers and many years for the interpreters.

With the inclusion of the teachers, we may start to tease apart the relation between individual differences in first and second language processes and cognitive capacity. If the differences between the students and the interpreters in Experiment 1 were solely based on differences in second language proficiency rather than interpreting experience, the teachers should perform very similarly to the interpreters. However, if the better performance of the interpreters is specifically related to their extensive interpreting experience, they should outperform the teachers.

Method

Participants

Fifteen teachers of English participated in this experiment, all of whom where native speakers of Dutch. They were all teaching at the higher levels of secondary education in the Netherlands. They had on average 19 years of experience in teaching English. The teachers were comparable to the interpreters in educational background (most participants held a Bachelors or Masters degree in English). They were also comparable in age. Participant characteristics and language questionnaire data are reported in Table 1.

Materials and procedure

The materials used in this experiment were the same as in Experiment 1. All tests were conducted on a Macintosh Powerbook G3 using the same apparatus and procedures as in Experiment 1. Nine teachers were tested in the laboratory, six were tested at their homes. The experiment was otherwise identical to Experiment 1.

Results

The same statistical analyses were conducted as in Experiment 1 except that the data for the teachers was compared to the data from the interpreters reported in Experiment 1.

Word retrieval

Table 10 gives the *F* statistics and degrees of freedom associated with each source in the analyses.

Picture naming

The mean correct picture naming reaction times for each group in the two languages are reported in Table 5. The omnibus repeated-measures analysis of variance with group (interpreters vs. teachers) as a between-subjects factor and language (Dutch vs. English) as within-subjects factor showed that there were no differences in latencies between groups (interpreters: 786 ms, teachers: 817 ms, 95% confidence interval = 85 ms). Similar to the results of Experiment 1, we found faster performance for Dutch than for English, 800–970 ms (95% confidence interval = 82 s). Like the interpreters and students in Experiment 1, the teachers were faster in Dutch than in English picture naming ($\Delta = 140$ ms; 95% confidence interval for the teachers: 58 ms).

For Dutch picture naming, the repeated-measures analysis of variance with group (interpreters vs. teachers) as a between-subjects factor and cognate status (cognates vs. noncognates) and frequency (high frequency words vs. low frequency words) as within-subject factors revealed no differences between latencies between teachers and interpreters, 819–781 ms (95% confidence interval = 115 ms). There were significant differences due to cognate status and frequency. Cognates were responded to faster than noncognates (746–854 ms, 95% confidence interval = 41 ms) and high frequency words faster than low frequency words (749–851 ms, 95% confidence interval = 38 ms). Furthermore, group, cognate status, and frequency interacted. For the teachers the cognate effect was smaller for high frequency words ($\Delta = 79$ ms) than low frequency words ($\Delta = 154$ ms), for the interpreters the effect was larger for or high frequency words ($\Delta = 139$ ms) than for low frequency words ($\Delta = 61$ ms). The 95% confidence interval for these differences was 51 ms for the teachers and 55 ms for the interpreters. The pattern of the cognate effects in the native language was therefore slightly different for the two groups.

It is noteworthy that for English picture naming we obtained no significant main effect of group, nor any interactions with this factor. As expected, we did obtain a significant differences due to cognate status and frequency, and these factors also interacted with each other. Cognates were responded to faster than noncognates (868–1073 ms, 95% confidence interval = 44 ms), and high frequency words faster than low frequency words (877–1063 ms, 95% confidence interval = 51 ms). The effect of cognates status was smaller for high fre-

Table 10
Analyses of variance results for Experiment: picture naming (Group × Language), picture naming Dutch (Group × Cognate status × Frequency), English (Group × Cognate status × Frequency), translation (Group × Language), translation English–Dutch (Group × Cognate status × Frequency), and Dutch–English (Group × Cognate status × Frequency)

Source	Degrees of freedom	F value	p
Picture naming (Group × Language)			
Group	1, 26	.02	.86
Language	1, 26	34.47***	<.001
Language × Group	1, 26	1.10	.30
Picture naming Dutch			
Group	1, 26	.48	.49
Cognate status	1, 26	28.52***	<.001
Cognate status × Group	1, 26	.17	.67
Frequency	1, 26	29.63	.53
Frequency × Group	1, 26	4.03	.06
Cognate status × Frequency	1, 26	.004	.95
Cognate status × Frequency × Group	1, 26	7.08	.21
Picture naming English			
Group	1, 26	.10	.75
Cognate status	1, 26	91.04***	<.001
Cognate status × Group	1, 26	.001	.97
Frequency	1, 26	56.12***	<.001
Frequency × Group	1, 26	1.14	.29
Cognate status × Frequency	1, 26	6.19*	.02
Cognate status × Frequency × Group	1, 26	1.91	.18
Translation (Group × Language)			
Group	1, 26	.73	.40
Language	1, 26	.01	.94
Language × Group	1, 26	.01	.95
Translation English–Dutch			
Group	1, 26	.63	.43
Cognate status	1, 26	165.64***	<.001
Cognate status × Group	1, 26	1.93	.18
Frequency	1, 26	203.54***	<.001
Frequency × Group	1, 26	.96	.33
Cognate status × Frequency	1, 26	49.67***	<.001
Cognate status × Frequency × Group	1, 26	.02	.87
Translation Dutch–English			
Group	1, 26	.56	.46
Cognate status	1, 26	85.37***	<.001
Cognate status × Group	1, 26	1.06	.31
Frequency	1, 26	80.80***	<.001
Frequency × Group	1, 26	2.36	.14
Cognate status × Frequency	1, 26	29.91***	<.001
Cognate status × Frequency × Group	1, 26	.59	.45

* $p < .05$.
*** $p < .001$.

quency ($\Delta = 150$ ms) than for low frequency words ($\Delta = 259$ ms). The 95% confidence interval for these differences was 47 ms.

Analyses of the percentage of errors showed that these did not significantly differ between the teachers and the interpreters, nor between the L1 and the L2 (teachers: 4.5 and interpreters: 5.3, 95% confidence interval = 1.16%, $F(1,26) < 1$, $p > .10$). There were more errors for English picture naming than Dutch picture naming (L1: 3.8%, and L2: 6.0%, 95% confidence interval = 1.11%, $F(1,26) = 8.77$, $p < .01$). In general, in the picture naming task we obtained a very similar pattern of results for the teachers and the interpreters.

Word translation

The mean correct word translation reaction times for each direction of translation are reported in Table 7. A repeated measures analysis of variance revealed no significant effect of group (interpreters: 786 ms, and teachers: 817 ms, 95% confidence interval = 82 ms) and no effect of translation direction. As was the case for the interpreters, for the teachers it did not matter whether they translated into or from the L1 (787–785, 95% confidence interval = 42 ms). This suggests that there were no differences in performance between the teachers and interpreters in word translation.

Separate analyses per language showed exactly the same pattern. Neither in Dutch–English translation nor in English–Dutch translation was there a significant effect of group, but significant effects on latency were obtained due to cognate status and frequency and a significant interaction between the latter two variables was obtained. For both translation directions cognates were responded to faster than noncognates (696–905 ms for Dutch–English translation, 95% confidence interval = 46 ms, and 701–903 ms for English–Dutch translation, 95% confidence interval = 28 ms) and high frequency words were responded to faster than low frequency words (717–883 ms for Dutch–English translation, 95% confidence interval = 46 ms, and 719–884 ms for English–Dutch translation, 95% confidence interval = 32 ms). For Dutch–English translation the effect of cognate status was smaller for high frequency words ($\Delta = 116$ ms) than for low frequency words ($\Delta = 301$ ms). The 95% confidence interval for these differences was 49 ms. Also for English–Dutch translation the effect of cognate status was smaller for high frequency words ($\Delta = 126$ ms) than for low frequency words ($\Delta = 278$ ms). The 95% confidence interval for these differences was 35 ms.

For the percentage of errors there was also no difference between groups (interpreters: 3.0%, teachers: 2.5%, 95% confidence interval = 0.7%, $F(1,26) < 1$, $p > .10$). More errors were made translating into English than into Dutch (2.0–3.6%, 95% confidence interval = 0.5%, $F(1,26) = 19.82$, $p < .001$).

Memory

The average performance on the memory tasks is given in Table 9. Table 11 gives the *F* statistics and degrees of freedom associated with each source in the analyses.

Reading span

A repeated-measures analysis of variance revealed a significant difference between groups. Performance was better for the interpreters than for the teachers 34.96–31.40, 95% confidence interval = 1.56. There was no significant difference between languages. Comparison against the 95% confidence interval for the teachers (1.45) showed that, similar to the results for the interpreters, the difference between Dutch and English reading span was not significant for the teachers ($\Delta = 1.33$). Furthermore, the 95% confidence interval for the difference between the means of the groups was 1.56. The interpreters performed significantly better than the teachers on both the Dutch ($\Delta = 3.32$), and the English reading span task ($\Delta = 3.81$).

Speaking span

For the speaking span task, analysis revealed a significant difference between groups (interpreters: 31.50, and teachers: 27.20, 95% confidence interval = 1.68) and languages (Dutch: 30.44, and English: 28.26, 95% confidence interval = .66), and an interaction between group and language.

In contrast to the interpreters, the teachers differed in performance on the Dutch and English versions of the task: Like the students in Experiment 1, the teachers performed significantly better on the Dutch version ($\Delta = 3.20$; 95% confidence interval for teachers = .91 ms). Furthermore, the 95% confidence interval for the difference between the means of the groups (1.68) suggests that the interpreters performed better than the teachers on both the Dutch ($\Delta = 3.28$) and the English version ($\Delta = 5.32$) of the speaking span task.

Table 11
Analyses of variance results Experiment 2: reading span (Group × Language), speaking span (Group × Language), and word span (Group × Language)

Source	Degrees of freedom	*F* value	*p*
Reading span			
Group	1, 26	10.92**	<.01
Language	1, 26	2.22	.15
Language × Group	1, 26	.11	.74
Speaking span			
Group	1, 26	13.81**	<.01
Language	1, 26	22.28***	<.001
Language × Group	1, 26	4.92*	.035
Word span			
Group	1, 26	13.32**	.001
Language	1, 26	.36	.55
Language × Group	1, 26	8.53**	.01

* $p < .05$.
** $p < .01$.
*** $p < .001$.

Word span

The analysis for the word span task revealed a significant difference between groups (interpreters: 5.46, and teachers: 3.1, 95% confidence interval = 1.56), and an interaction between group and language.

Again, in contrast to the interpreters, the teachers performed significantly better on the Dutch than on the English word span task ($\Delta = 1.40$; 95% confidence interval for teachers = .79). Furthermore, the 95% confidence interval for the difference between the means of the groups was 0.94. The interpreters performed better than the teachers on both the Dutch version ($\Delta = 3.52$) and on the English version ($\Delta = 1.20$) of the task.

Control tasks

On the vocabulary test and the arrow reaction time test independent samples *t* tests were conducted. These analyses showed that interpreters and the teachers performed similarly on both the vocabulary test, $t(26) < 1$, $p > .10$, and on the arrow test $t(26) = 1.49$, $p > .10$. The average performance of the teachers was .82 for the vocabulary test and 407 ms for the arrow test (the average performance was .86 and 418 ms, respectively, for the interpreters).

Summary of results

The teachers appeared to perform similarly to the interpreters on the lexical retrieval tasks. Moreover, on the two control tasks there was no difference between the teachers and the interpreters. This latter result suggests that our assumption was warranted that these two groups of participants were comparable to each other. The main conclusion we can draw from these results is that the interpreters were not more efficient in the retrieval of words than the teachers. Performance on the two processing tasks was virtually the same.

The critical finding in Experiment 2 was that the two groups performed differently on the memory tasks. Comparisons on the working memory span measures showed that the interpreters outperformed the teachers on all three tasks. The interpreters' performance was better regardless of the language of testing. In contrast to the interpreters, although the teachers were clearly very proficient in their L2, their performance on the memory tests was poorer when they were tested in L2 than in L1 (although not significantly so for the reading span test).

General discussion

The three participant groups in this study provided a unique opportunity to separate the effects of interpreting experience and language proficiency. The results of Experiments 1 and 2 may best be compared when the three participant groups are observed together. For this purpose, we plotted the data of the three groups for each of the tasks. Fig. 1 shows the results of the picture naming task and the word translation task, collapsed over word types.

These data show clearly that the interpreters and the teachers performed very similarly whereas the students were much slower (except in the only task that involves processing in the L1 exclusively, i.e., Dutch picture naming). We may conclude from this that efficient lexical retrieval is not uniquely relevant for interpreting and may be mediated by general language proficiency. This does not imply that fluent word retrieval is not important in simultaneous interpreting. In fact, we recently observed a high correlation between word retrieval skill and interpreting performance in participants without any previous experience in simultaneous interpreting (Christoffels et al., 2003). The present finding that the teachers performed similarly to the interpreters merely indicates that word retrieval is not uniquely related to simultaneous interpreting. Lexical retrieval is not 'boosted' any further by professional interpreting than by another profession that demands high proficiency in the L2 (i.e., the teaching of English). This latter finding illustrates that, when searching for specific subskills involved in interpreting, one has to be careful in choosing an appropriate comparison group. Note, however, that the latencies in English picture naming and translation for both the interpreters and the teachers were shorter than those of the students even though the arrow test indicated that the basic, nonlinguistic reaction time of the teachers and interpreters was slower.

In the working memory tasks, the pattern of results was clearly different. The data for the reading span, speaking span, and word span tasks for the three groups are presented in Fig. 2.

This figure shows that in general the interpreters not only had a higher memory capacity than the students and the teachers, but also that they performed similarly in L1 and L2. In contrast to the lexical retrieval tasks, on this set of tasks the teachers' performance was similar to that of the students, not to that of the interpreters. Like the students, the teachers show an effect of language: their memory capacity was larger when they were tested in L1. We will discuss this language effect later.

In sum, the combined results of Experiments 1 and 2 showed that the teachers performed like the interpreters on the word retrieval tasks and on the control tasks, but like the bilingual students on the working memory tasks. To further examine this result, we conducted an additional set of analyses to compare the bilingual students directly to the teachers. The analyses on the lexical retrieval tasks produced the same pattern of results as the analyses comparing the students to the interpreters. The teachers, like the interpreters, were faster to perform these word production tasks than the bilingual students. However, on the memory span tasks the teachers and the students did not differ from each other.[3] Thus, it seems that professional interpreting experience is specifically associated with elevated verbal memory capacity. The relation of skilled interpreting to superior memory skills is perhaps even more compelling if we consider that the teachers and the students either have been, or are currently, enrolled in university-level education, which also involves frequent use of English. The teachers and students can therefore be thought to represent the high end of the scale both in terms of proficiency and capacity, suggesting that the interpreters' performance on the memory tasks is truly exceptional.

Recent research provides evidence that memory capacity and word translation performance may be related, in the sense that higher cognitive capacity is associated with better performance on a word translation task (Kroll et al., 2002). Nevertheless, a graphical models analysis indicated that working memory tasks and translation each had an independent effect on simultaneous interpreting-performance (Christoffels et al., 2003). Furthermore, the present study showed a dissociation between the two subsets of tasks, working memory and lexical retrieval. The fact that, on the one hand, the teachers may be distinguished from the students by their performance on the lexical retrieval tasks, and on the other hand, from the interpreters by their performance on the working memory tasks suggests that lexical retrieval and working memory may be regarded as different independent dimensions of participants' skills.

[3] Repeated-measures analysis of variance on the picture naming task involving the students and the teachers as levels of the factor group revealed marginally significant main effects of group, indicating that the teachers were faster than the bilinguals, $F(1,50) = 3.32$, $p = .074$, $\eta^2 = .06$, and a significant interaction between group and language, $F(1,50) = 16.13$, $p < .001$, $\eta^2 = .24$. For the word translation task the effect of was group also significant, $F(1,50) = 16.28$, $p < .001$, $\eta^2 = .24$. In contrast, similar analyses on the working memory tasks revealed no effects of groups, reading span $F = 1.34$; speaking span and word span, Fs < 1, but the main effects of language were significant, indicating better performance on the Dutch version of the tasks, reading span $F(1,52) = 11.05$, $p = .001$; $\eta^2 = .18$; speaking span, $F(1,52) = 42.42$, $p < .001$, $\eta^2 = .45$; word span, $F(1,52) = 11.55$, $p = .001$, $\eta^2 = .18$.

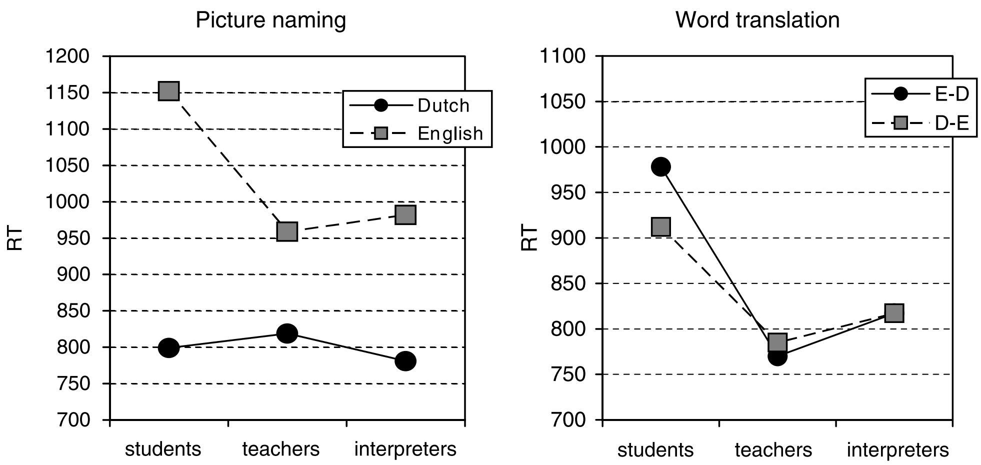

Fig. 1. Average reaction times on the picture naming and word translation tasks per group.

Bilingual working memory

The superior short-term and working memory skills of the interpreters may be related to several task components of simultaneous interpreting. As mentioned in the introduction, interpreting may be regarded as a dual-task situation because of the simultaneity of comprehension and production processes. Short-term memory is taxed because of the time-lag between input and output. Furthermore, the requirement to manage two languages may be related to increased cognitive control. The interpreters were better at both the working memory and short-term memory tasks. Storage and processing aspects of (working) memory have been shown to have mutual influence in a reading span task (Saito & Miyake, 2004) suggesting that both storage and processing aspects are important. In this context it is noteworthy that interpreters are resistant to the detrimental effects of articulatory suppression (e.g., Padilla et al., 2005) and that recall under these conditions is related to interpreting performance in untrained interpreting (Christoffels, in press). Perhaps one aspect of efficiency in processing may be the efficiency of transfer of information from a short-term memory presentation to a semantically based longer-term memory (see Christoffels, in press).

In the present study, memory capacity was assessed in both the L1 and the L2, because functional memory capacity may differ in two languages. Indeed, for two of our participant groups, the teachers and the students, it did. In general, a straightforward interpretation of the differences between native and nonnative span is that they are caused by the fact that processing in the L2 is less efficient and therefore takes up more resources, leaving less capacity available for storage (e.g., Just & Carpenter, 1992; Bayliss, Jarrold, Gunn, & Baddeley, 2003). Although this is bound to be an important factor, we also found recall differences in the two languages on the word span task, a task that does not require both the storage and processing of information. Clearly, there is some difference between our groups even in the ability to passively store verbal information in the second language.

We used two different types of working memory tasks. Daneman and Green (1986) argued that resources may be domain specific, and that therefore different processing tasks should be used when investigating the role of capacity in different domains (i.e., language comprehension versus production). On the other hand, Daneman and Merikle (1996) compared verbal processing tasks with mathematical processing tasks, and concluded that it is not important how exactly verbal working memory is measured to be predictive of language comprehension. The question whether or not resources are domain specific is beyond the scope of this paper (see MacDonald & Christiansen, 2002; for a discussion of this issue). Nevertheless, since we assessed capacity in two languages, the question presents itself whether the particular task we used or the language we tested in was the more important determinant of performance. To examine this issue, the correlations between the different versions of the memory test may be informative. They showed that nearly all versions of all memory tasks correlated significantly with each other, suggesting that they tap into a common factor. The correlations were not systematically higher between the Dutch and the English version of the same task than between the same language versions of different tasks (Dutch vs. English reading span, $r = .45$; Dutch vs. English speaking span, $r = .63$; reading span vs. speaking span Dutch, $r = .55$; reading span vs. speaking span English, $r = .51$, p's $< .001$), but the latter correlations tended to be higher than when both task and language differ. This suggests that task type and language of testing are equally important in determining performance. Correlations of

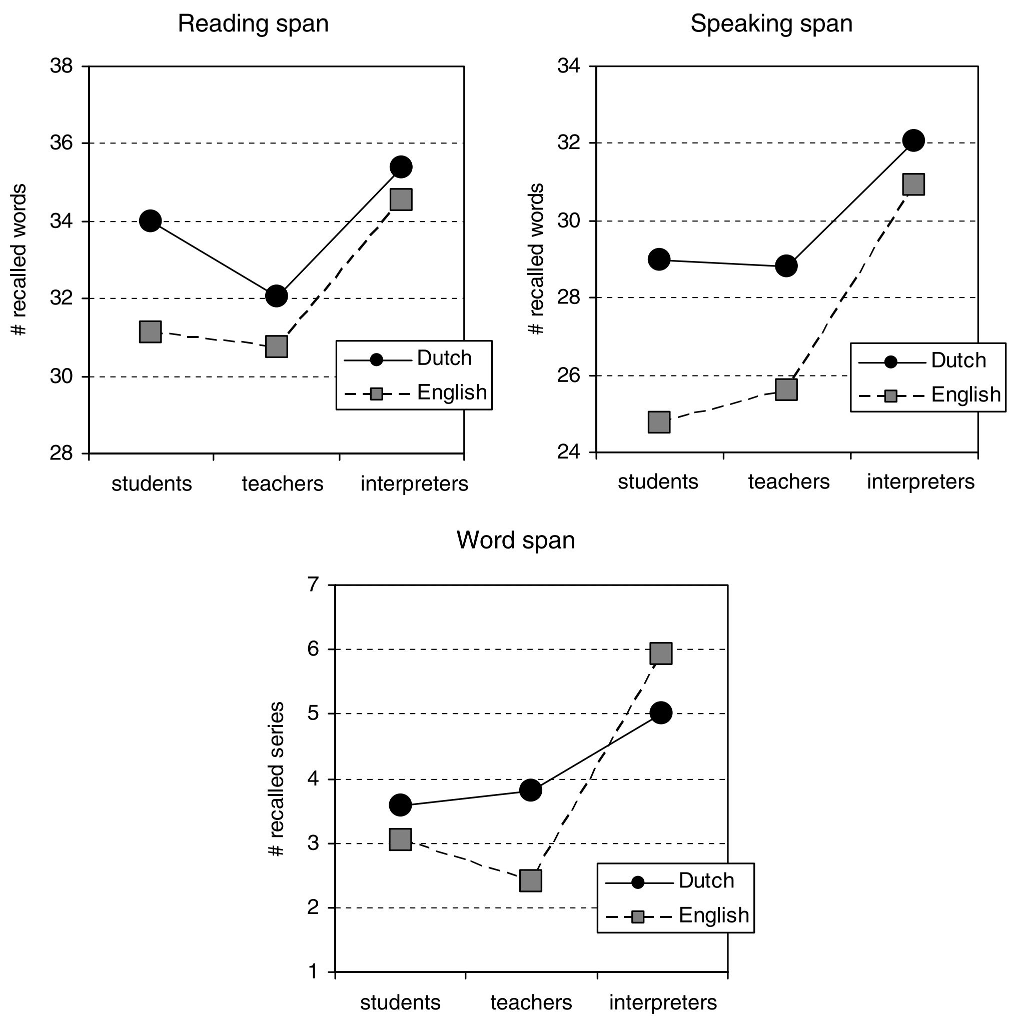

Fig. 2. Average scores on the reading span, speaking span, and word span tasks per group.

the working memory tasks with the word span task were somewhat lower, especially in Dutch (Dutch vs. English word span: $r = .53$, reading span vs. word span in Dutch: $r = .28$, reading span vs. word span English: $r = .44$; speaking span vs. word span Dutch $r = .28$, speaking span vs. word span English $r = .49$, p's $< .05$).

A rather intriguing finding is the fact that the teachers, unlike the interpreters, showed a language effect in the memory tasks even though they performed very similar to the interpreters on all language related tasks (i.e., the two lexical retrieval tasks and the English vocabulary test). Especially in the L2 the teachers' functional capacity was smaller than the interpreters' capacity. It seems that although the knowledge of English of the two groups is very similar, the interpreters' processing and storage in English was more efficient. In other words, the two groups are differentially affected by language in a way that manifests itself in processing and storage capacity rather than in the speed of lexical retrieval.

Frequency and cognate effects

Pictures with high-frequency names were named faster than pictures with low frequency names, and similarly, high frequency words were translated faster than low frequency words, replicating previously reported effects of word frequency (e.g., De Groot, 1992; Jescheniak & Levelt, 1994). We also found strong effects of cognate status, in both word translation and picture naming. Cognates have been found to facilitate translation (e.g., De Groot et al., 1994; Kroll et al., 2002; Kroll & Stewart, 1994; Sanchez Casas et al., 1992) and picture naming (Christoffels et al., 2003; Costa et al., 2000; Kroll, Dijkstra, Janssen, & Schriefers, in preparation).

Thus, we replicated these effects for the students and extended them to teachers and interpreters, groups that, to our knowledge, have not been tested before on these tasks.

In translation, the size of the cognate-effect has been shown to depend on proficiency in the L2 and also on memory span for second language learners at early stages of acquisition (Kroll et al., 2002). Therefore, differences in L2 proficiency and interpreting experience in our participants may result in differences in how cognate status influences latency on the word retrieval tasks. The interpreters especially may be hypothesized to exercise strong control on the relative activation of information in each of their languages. If the interpreters are better able to modulate the influence of the nontarget language, the cognate effect may be attenuated. To see whether this was the case, we performed a separate analysis with the factor group as between-subject factor (with all three participant groups), and language and cognate status as within-subject-factors (i.e., we collapsed over frequency). The interaction between group and cognate status did not reach significance, neither for picture naming nor for translation ($F < 1$ and $F = 1.01$, respectively). This strongly suggests that there are no qualitative differences between groups with respect to the cognate effect. It seems then that the magnitude of this effect is not influenced by the amount of processing resources available or by language proficiency in our three relatively proficient participant groups.

Interestingly, for each of the groups we obtained a cognate effect in the Dutch picture naming task (see results sections). In other words, even though no words were actually presented in the picture naming task, the fact that the name of the picture in L1 has a translation equivalent in the L2 that is similar in form influenced latencies in this L1 only task. As mentioned earlier, there is a great deal of support for language nonselectivity in word recognition. Many past word recognition studies have shown an effect of the L1 on L2 performance (e.g., Bijeljac Babic, Biardeau, & Grainger, 1997; De Groot et al., 2000; Dijkstra et al., 1998; Van Heuven et al., 1998) and some studies have shown an effect of L2 on L1 (e.g., Jared & Kroll, 2001; Van Wijnendaele & Brysbaert, 2002). Van Hell and Dijkstra (2002) recently showed that participants from the same population as our students and tested in a strict monolingual setting responded faster to L1 words that were cognates than to noncognates in lexical decision and word-association tasks.

Our findings extend those of Costa et al. (2000), who observed a facilitating effect for cognates in picture naming for Spanish–Catalan bilinguals. They found that the effect of cognate status was larger when naming in the nondominant language than when naming in the dominant language (L1). We similarly obtained an interaction between language and cognate status, ($F(1,64) = 21.01$, $p < .001$), which was larger for picture naming in the L2 ($\Delta = 201$ ms) than in the L1 ($\Delta = 100$ ms). The cognate facilitation effect suggests that not only L2 but even L1 language production is nonselective with respect to language. The L1 language system is shown to be influenced even in groups who are not as balanced as the Spanish–Catalan population of Costa et al. (2000).

The pattern of findings has interesting implications for models of bilingual processing and language control. We showed that interpreters, who given the nature of their profession should be able to exert strong control of their language output and who demonstrated strong working memory skills, nevertheless show the same cognate facilitation effect as the other groups in this study. In translation the stimulus is presented in the nontarget language, therefore an effect of the relation between the stimulus and the target word might not seem surprising. However, the picture naming data suggests that the phonology, and perhaps the orthography, of the nontarget language is activated automatically, beyond the influence of control exerted by the interpreters.

The presence of a cognate effect in picture naming for the interpreters despite their very high memory capacity suggests that, contrary to recent claims about the allocation of cognitive resources in modulating cross-language activity (e.g., Green, 1998; Michael & Gollan, 2005), it may not be possible to selectively inhibit the non-target language, at least not in this sort of out of context task. How then can language of output be controlled? Several authors proposed that language membership may somehow be used for selective lexical activation (e.g.,La Heij, 2005), for language specific lexical selection (e.g., Costa, 2005; Costa et al., 2000), or for a checking mechanism in the inhibitory-control account of Green (1998). It seems that our data are in accordance with the latter type of control rather than with a view of global inhibition of the nontarget language (e.g., Green, 1998; Paradis, 1994; see also Christoffels & De Groot, 2005, for a discussion of language control in simultaneous interpreting). In future studies it will be of interest to address possible other cognitive abilities such as cognitive control further by focusing on these abilities, using for example language switching and conflict tasks (e.g., Meuter & Allport, 1999; Bialystok et al., 2005).

To conclude, the main result of the experiments reported in this paper is that the memory performance of the interpreters was better than that of both the students and the teachers. Furthermore, lexical retrieval, as measured by on-line processing tasks, was not more efficient for the interpreters than for the teachers, who are also highly proficient in their L2. Interestingly, whereas the interpreters' working memory and short-term memory capacity was not influenced by language of testing, the capacity for the other groups was superior in the L1. This result suggests that the interpreters have developed greater efficiency in language processing capacity in the L2 than ordinary bilinguals. The data strongly suggest that specifically working memory is an important subskill for simultaneous interpreting whereas the role of lexical retrieval may be mediated by general language proficiency. In all, we showed that a particular kind of bilingual expertise, simultaneous interpreting, is selectively associated with enhanced memory capacity in both the native and the second language. In future research it will be important to learn whether this enhanced memory is a prerequisite or a consequence of simultaneous interpreting experience and, more generally, to understand the role of working memory in language processing in context.

References

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

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.

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

Bialystok, E. (2005). Consequences of bilingualism for cognitive development. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of Bilingualism: Psycholinguistic Approaches* (pp. 417–432). New York: Oxford University Press.

Bialystok, E., Craik, F. I. M., Grady, C., Chau, W., Ishii, R., Gunji, A., et al. (2005). Effect of bilingualism on cognitive control in the Simon task: Evidence from MEG. *NeuroImage, 24*, 40–49.

Bijeljac Babic, R., Biardeau, A., & Grainger, J. (1997). Masked orthographic priming in bilingual word recognition. *Memory & Cognition, 25*, 447–457.

Chen, H. C., & Leung, Y. S. (1989). Patterns of lexical processing in a nonnative language. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 316–325.

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. (2004). *Cognitive studies in simultaneous interpreting.* Unpublished doctoral dissertation, University of Amsterdam, Amsterdam.

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

Christoffels, I. K., & De Groot, A. M. B (2004). Components of simultaneous interpreting: A comparison with shadowing and paraphrasing. *Bilingualism: Language and Cognition, 7*, 1–14.

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. 326–348). 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 lexical retrieval, working memory, and simultaneous interpreting. *Bilingualism: Language and Cognition, 6*, 201–211.

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

Costa, A. (2005). Lexical access in bilingual production. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 308–325). New York: Oxford University Press.

Costa, A., Caramazza, A., & Sebastian-Galles, N. (2000). The cognate facilitation effect: Implications for models of lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 1283–1296.

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.

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

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

De Groot, A. M. B. (1992). Determinants of word translation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*, 1001–1018.

De Groot, A. M. B. & Christoffels, I. K. (in press). Language control in bilinguals: Monolingual tasks and simultaneous interpreting. *Bilingualism: Language and Cognition*.

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., Delmaar, P., & Lupker, S. J. (2000). The processing of interlexical homographs in translation recognition and lexical decision: Support for non-selective access to bilingual memory. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 53A*, 397–428.

De Groot, A. M. B., & Nas, G. L. (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 of proficiency in a second language: The ubiquitous involvement of conceptual memory. *Language Learning, 47*, 215–264.

Dijkstra, T., Van Jaarsveld, H., & Ten Brinke, S. (1998). Interlingual homograph recognition: Effects of task demands and language intermixing. *Bilingualism: Language and Cognition, 1*, 51–66.

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

Garrod, S., & Pickering, M. J. (2004). Why is conversation so easy. *Trends in Cognitive Sciences, 8*, 8–11.

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

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 J. H. Danks, G. M. Shreve, S. B. Fountain, & M. K. McBeath (Eds.), *Cognitive processes in translation and interpreting* (pp. 196–214). Thousand Oaks: Sage Publications.

Goldman-Eisler, F. (1972). Segmentation of input in simultaneous translation. *Journal of Psycholinguistic Research, 7*, 127–140.

Green, D. W. (1986). Control, activation, and resource: A framework and a model for the control of speech in bilinguals. *Brain and Language, 27*, 210–223.

Green, D. W. (1998). Mental control of the bilingual lexico-semantic system. *Bilingualism: Language and Cognition, 7*, 67–81.

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., Pickering, M. J., & Veltkamp, E. (2004). Is syntax separate or shared between languages? Cross-linguistic syntactic priming in Spanish-English bilinguals. *Psychological Science, 75*, 409–414.

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

Hermans, D., Bongaerts, T., De Bot, K., & Schreuder, R. (1998). Producing words in a foreign language: Can speakers prevent interference from their first language? *Bilingualism: Language and Cognition, 1*, 213–229.

Huibregtse, I., & Admiraals, W. (1999). De score op een ja/nee-woordenschattoets: Correctie voor raden en persoonlijke antwoordstijl. [The score on a yes/no vocabulary test: Correction for guessing and individual style of answering]. *Tijdschrift voor Onderwijsresearch, 24*, 110–124.

Jared, D., & Kroll, J. F. (2001). Do bilinguals activate phonological representations in one or both of their languages when naming words. *Journal of Memory and Language, 44*, 2–31.

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*, 824–843.

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., & De Groot, A. M. B. (1997). Lexical and conceptual memory in the bilingual: Mapping form to meaning in two languages. In A. M. B. de Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 169–199). Mahwah: Lawrence Erlbaum Associates.

Kroll, J. F., Dijkstra, A., Janssen, N., & Schriefers, H. (in preparation). *Selecting the language in which to speak: Cued-picture naming experiments on lexical access in bilingual production*.

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. (2005). Selection processes in monolingual and bilingual lexical access. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 289–307). New York: Oxford University Press.

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. (1992). Shadowing. *Méta, 37*, 263–273.

Loebell, H., & Bock, K. (2003). Structural priming across languages. *Linguistics, 41*, 791–824.

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review, 1*, 476–490.

MacDonald, M. C., & Christiansen, M. H. (2002). Reassessing working memory: A comment on Just & Carpenter (1992) and Waters & Caplan (1996). *Psychological Review, 109*, 35–54.

Macizo, P., & Bajo, M. T. (in press). Reading for repetition and reading for translation: Do they involve the same processes? *Cognition*.

Masson, M. E. J., & Loftus, G. R. (2003). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology, 57*, 203–220.

Meara, P. (1996). *English vocabulary test 10K*. Swansea, UK: Centre for Applied Language Studies, University of Wales.

Meuter, R. F. I., & Allport, A. (1999). Bilingual language switching in naming: Assymetrical costs of language selection. *Journal of Memory and Language, 40*, 25–40.

Michael, E. B., & Gollan, T. (2005). Being and becoming bilingual: Individual differences and consequences for language production. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 389–407). New York: Oxford University Press.

Miller, N. A., & Kroll, J. F. (2002). Stroop effects in bilingual translation. *Memory & Cognition, 30*, 614–628.

Miyake, A. (1998). Individual differences in second language proficiency: The role of working memory. In A. F. Healy & L. E. Bourne, Jr. (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 339–364). Mahwah, NJ: Lawrence Erlbaum Associates.

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* (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.

Potter, M. C., So, K. F., Von Eckardt, B., & Feldman, L. B. (1984). Lexical and contextual representation in beginning and more proficient bilinguals. *Journal of Verbal Learning and Verbal Behavior, 23*, 23–28.

Saito, S., & Miyake, A. (2004). On the nature of forgetting and the processing-storage relationship in reading span performance. *Journal of Memory and Language, 50*, 425–443.

Sanchez Casas, R. M., Davis, C. W., & Garcia Albea, J. E. (1992). Bilingual lexical processing: Exploring the cognate/non-cognate distinction. *European Journal of Cognitive Psychology, 4*, 293–310.

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.

Smith, M. C. (1997). How do bilinguals access lexical information. In A. M. B. de Groot & J. F. Kroll (Eds.), *Tutorials in bilingualism: Psycholinguistic perspectives* (pp. 145–168). Mahwah, NJy: Lawrence Erlbaum Associates.

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. (1998a). Conceptual representation in bilingual memory: Effects of concreteness and cognate status in word association. *Bilingualism: Language and Cognition, 1*, 193–211.

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

Van Hell, J. G., & Dijkstra, T. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic Bulletin & Review, 9*, 780–789.

Van Heuven, W. J. B., Dijkstra, T., & Grainger, J. (1998). Orthographic neighborhood effects in bilingual word recognition. *Journal of Memory and Language, 39*, 458–483.

Van Wijnendaele, I., & Brysbaert, M. (2002). Visual word recognition in bilinguals: Phonological priming from the second to the first language. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 616–627.

Zeelenberg, R., & Pecher, D. (2003). Evidence for long-term cross-language priming in conceptual implicit memory tasks. *Journal of Memory and Language, 49*, 80–94.