Working memory performance in expert and novice interpreters

Barbara Köpke, Jean Luc Nespoulous

Institutions: University of Toulouse

Published on: 01 Jan 2006 - Interpreting (John Benjamins)

Topics: Articulatory suppression, Free recall, Recall, Working memory and Stroop effect

Working memory performance in expert and novice interpreters

Barbara Köpke, Jean-Luc Nespoulous

Working memory performance in expert and novice interpreters.

Barbara Köpke & Jean-Luc Nespoulous

Laboratoire Jacques Lordat
Institut des Sciences du Cerveau de Toulouse
Université de Toulouse-Le Mirail
5, allées Antonio Machado
31058 Toulouse Cedex
France

bkopke@univ-tlse2.fr
nespoulo@univ-tlse2.fr

Abstract (203 words)
Simultaneous interpreting is generally assumed to be particularly demanding with respect to cognitive resources like attention and working memory, which are thought to gradually increase with professional practice. Experimental data to corroborate such an assumption is still sparse, however. Here we report an in-depth investigation of working memory capacity among 21 professional interpreters (experts), 18 second-year interpreting students (novices) and two control groups (20 multilinguals and 20 students). Tests involved either short-term retention alone; short term retention and processing in a recall task with articulatory suppression, a listening span task, and a category and rhyme probe task; or attention alone in a unilingual and bilingual Stroop test. No between-group differences in simple span tasks and the Stroop test were found. Significant group effects were observed in free recall with articulatory suppression, in the category probe task and in the listening span task. The best performance was always produced by the novice interpreters rather than by the experts. These findings are discussed in relation to a) the novice-expert distinction and the role of working memory in the development of interpreting skills, and b) the nature of the task and possible strategies involved.

Keywords: simultaneous interpreting, working memory, expertise, capacity view, skill development, semantic short-term memory, phonological short-term memory

Short biographical sketch:

Barbara Köpke received her PhD in Linguistics from the University Toulouse-Le Mirail where she is currently Associate Professor. Her main research interests concern language processing in bilingual subjects with special focus on 'extreme situations' such as simultaneous interpreting and attrition. She conducted this research program as a 'Young Fellow' from the French Ministry of Research within the context of a Grant LACO 47 from the "Cognitique" Program given to the second author of the paper.

Jean-Luc Nespoulous is Professor of neuropsycholinguistics in the University of Toulouse-Le Mirail. He is currently Director of both the Jacques-Lordat Laboratory (E.A. 1941) and the Toulouse Brain Science Institute (IFR 96). He has been mainly involved in research on language pathology over the past 30 years.

Introduction

While it has long been assumed that cognitive factors such as memory play an important role in such a highly complex task as simultaneous interpreting, experimental research into these factors has been sparse (e.g. Chincotta & Underwood 1998; Darò & Fabbro 1994; Padilla Benítez 1995). A major event in the development of a theoretical frame for research into the cognitive aspects of interpreting was the Second Ascona Conference (2000) which was aimed at establishing the study of interpreting as a research paradigm within the study of complex cognitive processes (see http://mambo.ucsc.edu/ascona/announcement.html), and gave rise to a number of theoretical papers exploring the links between various aspects of cognitive theory and interpreting. Theoretical models that have received specific attention within this context are for example immediate memory (Hulme 2000), working memory following the different versions of Baddeley's model (Baddeley 1986; 2000a; 2000b) or Cowan's focus of attention (Cowan 2000a; 2000b) as well as the Ericsson & Kintsch framework of longterm working memory (e.g. Ericsson 2000; Ericsson & Kintsch 1995).

The complexity of simultaneous interpreting need not be outlined further (see for example Cowan 2000; Frauenfelder & Schriefers 1997; Moser-Mercer 2000; Pöchhacker 2004). The highly legitimate hypothesis that memory and attention skills play a major role in the process of interpreting has been discussed by many authors and interesting predictions have been made (e.g. Cowan 2000; Darò & Fabbro 1994; Hulme 2000). The few experimental data, however, are still far from any conclusion.

One of the first studies to have focused on short term retention skills in simultaneous interpreting is the study of Darò & Fabbro (1994) demonstrating the cognitive complexity of the interpreting task. The authors asked 24 beginner interpreters (students with around 2 years of professional experience) to perform a digit span task in 4 different conditions: after listening, shadowing, articulatory suppression and simultaneous interpretation. The results showed that the subjects' performance in the digit span task was significantly lower after simultaneous interpreting compared to the other conditions. This result was taken as evidence that interpreting is the most complex task, disrupting performance on a concurrent task to an important degree due to phonological interference. This conclusion led to the assumption that the professional practise of simultaneous interpreting would develop the ability to resist to phonological interference in short term memory tasks and in particular in tasks where cognitive demands are enhanced through phonological interference (such as in articulatory suppression for example).

Together with the assumption that the simultaneous interpreting task, by its very nature, puts high demands on memory and attention skills (e.g. Hulme 2000), a number of studies were undertaken aimed specifically at testing working memory skills in professional interpreters. The rationale of these studies involves the idea that such skills develop during training, so most of them compare the memory performance of novice and expert interpreters, i.e. students and professionals. Some also include an external control group. Since methodological issues have not received much attention, we will present these studies in detail here (Chincotta & Underwood 1998; Liu, Schallert & Carroll 2004; Nordet & Voegtlin 1998; Padilla Benítez 1995).

Experimental evidence

One of the most frequently used tasks for the testing of interpreter's cognitive capacity is some variant of the reading span or listening span task introduced by Daneman & Carpenter (1980). This paradigm indeed appears particularly interesting in this context since it has been linked to higher individual cognitive capacity, particularly regarding comprehension skills (Just & Carpenter 1992).

Reading or listening span

One of the first studies of interpreting based on this paradigm was the dissertation from Padilla Benítez (1995). The task was a reading span adapted to Spanish, the first language of all subjects: they had to read aloud sentences presented visually and remember the last word of each sentence. Four groups of participants were tested: 10 professionals, 10 control subjects with similar education level, 10 advanced level interpreting students and 10 beginning level interpreting students. It should be noted that the professionals had only little professional experience since half of them actually had no experience (they had just obtained the diploma) and the other half of the participants had between 4 and 5 years of experience. They were aged from 23 to 33 years. The control subjects all had a university degree in the humanities and were between 25 and 35 years old. Findings indicate that the reading span was significantly higher in the professionals than in all other groups and there was no difference between the control group and both student groups.

It can be argued that the reading span is not the most relevant measure with respect to simultaneous interpreting since it is based on visual material, whereas interpreting can be supposed to involve auditory memory rather than visual memory skills. For this reason, Nordet & Voegtlin (1998) developed an auditory version of the reading span, a listening span where sentences were presented aurally and subjects had to perform a veracity judgment while memorizing the last word. In addition, the authors varied some other parameters such as sentence length, final word length and concreteness. The participants were 6 professional interpreters with a mean age of 35,8 years, 7 interpreting students (including one visually disabled student) with a mean age of 26,3 years, and a control group of 22 students in psychology (mean age = 22,6 years). The results show no significant differences between the groups. Note however that there was a tendency for the interpreting students to be better than the professionals, both groups performing slightly better than the control students.

The last study in which such a working memory measurement has been used is, as far as we know, the recent study from Liu, Schallert & Carroll (2004) where a listening span was actually used as a measure of general cognitive ability to be compared with performance on an interpreting task. The listening span implied a judgment of veracity and the recall of the final words and was carried out in English, the second language of most of the subjects. Participants were 11 professionals with at least 2 years of experience, 11 beginners (first year students) and 11 advanced (second year students). The results showed no differences between groups in the listening span, but differences between interpreters and students in interpreting measures: in the semantic tasks associated to the interpreting task, professionals clearly scored higher than the students. The authors conclude that the better performance of the experts is actually due to their professional experience and not to greater cognitive ability as measured by the listening span.

Digit span

Besides the Darò & Fabbro (1994) study—in which the digit span measure actually was not used in order to investigate cognitive ability in interpreters, but as evidence of the high cognitive demands of the interpreting task—a digit span task was used in two further studies. Padilla Benítez (1995), in the study described above, presented digit lists auditorily according to the usual procedure. She found that professionals were better than the other participants of the study. Chincotta & Underwood (1998) used a digit span task with visual presentation in order to investigate the effect of articulatory suppression on short term memory performance. Participants in the study were 12 professional Finnish-English interpreters with a minimum of 100 hours experience and a control group of 12 undergraduate students of English. Digit span measures in Finnish and English were compared to reading time for digits in order to test whether the correlation between reading time and digit span would be eliminated by articulatory suppression. Findings showed that articulatory suppression entailed lower performance for both groups. The bilingual digit span effect, however, was abolished by articulatory suppression in the control group only and persisted in the professional interpreters. There were, however, no differences between groups with respect to digit span alone.

Free immediate recall

Padilla Benítez (1995) further tested her subjects (described above) in immediate free recall with and without articulatory suppression (there was also a measure of delayed recall of the same material that will not be reported on here). The material consisted of 3 lists of 16 words for each condition, presented visually at intervals of 3 seconds. The findings showed that professionals were significantly better than all other groups in the articulatory suppression condition but not in the control condition. Again, there was no difference between interpreting students and controls.

The use of free recall tasks in this context has been questioned by other authors (e.g. Chincotta & Underwood 1998: 7). Indeed, a free recall task of this kind probably exceeds phonological rehearsal capacity both in terms of the number of items (16 items per list) and in terms of a time limit (16 items presented at 3s intervals = 48 s). The task therefore is most likely to rely also on central executive components. This however renders the task very promising in the context of research with interpreters. Additionally, the constraint to recall items in the original order is an important constraint which might not have its validity in many real life memory tasks. In many cases, the order of items is either not relevant, or it can be inferred from the items (as in most cases of sentence processing). A free recall task might hence be specifically indicated for exploring the interpreter's capacity to process and organise the material to be memorized.

On the whole, the few studies investigating working memory in expert and novice interpreters allow for only very tentative conclusions. All studies involved small numbers of participants. The only study finding evidence for higher memory skills in professionals was the Padilla Benítez (1995) study. This finding is, however, not confirmed by the other studies and it cannot be excluded that it is due to characteristics of the experimental groups (cf. below). Furthermore, comparison between the studies is delicate since in each study different groups of participants were compared. The only study comprising a control group matched to the professionals was the Padilla study; in the other studies, the control group—when there was one—consisted of undergraduate students who differed in both age and education level from the professionals.

Furthermore, these studies presented only very partial measures of working memory. Following the results of the Darò & Fabbro study, most studies started from the assumption of phonological interference caused by simultaneous listening and speaking and focused on the phonological loop. Findings from neuropsychological studies, however, suggest that short-term retention implies separate capacities for phonological and semantic retention. For instance, it has been shown that differential impairment of phonological and semantic aspects of language in aphasics has consequences on the retention of linguistic material in short term memory (Martin & Saffran 1997). In other cases, dissociations between semantic and phonological retention deficits have been described (Martin, Shelton & Yaffee 1994) supporting a multi-capacity view of short term memory. It is assumed here that a capacity that can be selectively disturbed in pathology might be particularly developed in the case of experts. Surprisingly enough, semantic aspects of working memory in the context of simultaneous interpreting have not yet received any attention.

On account of this lack of experimental evidence, we consider the discussion on both short term and working memory in the context of interpreting to be still open. The aim of the present study was to investigate this question by means of a study based not only on a statistically valid number of participants but also on a rigorously controlled methodology. The study is meant to give a broad survey of a variety of memory tasks—tapping into more peripheral as well as central aspects of working memory—in novice and expert interpreters with special attention to both semantic and phonological capacities.

Method

Participants

A total of 79 subjects participated in the study, 39 experimental subjects (21 professional interpreters and 18 interpreting students) and 40 control subjects.

- The interpreters (*experts*) were all highly skilled professionals working either on a permanent (N=12) or on a free-lance basis (N=9) for international institutions in either Brussels or Paris. The A language of all interpreters was French and all of them had English (among others) as either second A, B, or C language. Professional experience ranged from 4 to 35 years with a mean of 16,9 years. Age ranged from 29 to 61 years with a mean of 44,4 years.
- The interpreting students (*novices*) were all second (=last) year students, which implies that they had just started training in simultaneous interpreting[i]. They attended three different interpreting schools in either Brussels or Paris. They all had French as A language and English (among others) as either B or C language. Age ranged from 23 to 38 years, with a mean age of 26,2.

These experimental subjects were compared to two control groups.

- The first (*bilinguals*) consisted of 20 French-English bilinguals (or multilinguals) with at least 5 years of higher education. All of them had spent several years in an English-speaking country, but none had ever worked as a conference interpreter. Many of them were teaching languages in either secondary or higher education, others were engineers, economists, translators or researchers. Age was controlled for in relation to the interpreters group: it ranged from 27 to 63 years with a mean of 44,7 years.
- The second control group (*students*) consisted of 20 university students, all French native speakers without any particular competence regarding other languages. Their ages ranged from 18 to 26 years with a mean of 21,5 years.

Materials

Except for the listening span (which was based on the material conceived by Desmette et al. 1995) all memory tasks were constructed on the basis of a list of 738 selected French words and 225 pseudo words respecting French phonological constraints. The words were all common nouns controlled with respect to frequency, length and concreteness which are known to influence short term memory performance (e.g. Hulme, Roodenrys, Schweickert, Brown, Martin & Stuart 1997; Walker & Hulme 1999). Following the possibilities given by the French language, all words were low frequency[ii], bi-syllabic, concrete nouns. Furthermore, in order to fulfill the requirements of the phonologically constrained tasks, all words started with a consonant. The pseudo words were constructed by changing one or both vowels of 225 of these words (see examples below). Except for the English version of the Stroop Test, all tasks were conducted in French.

Task design

The participants responded to three types of tasks: (a) tasks involving storage only aimed at measuring short term retention in the phonological loop (b) tasks involving storage and processing, tapping more into the central executive (c) finally, in order to dissociate storage from attention and since such measures have never been applied to simultaneous interpreting, we added a Stroop test.

(a) Tasks involving storage only:
This group of tasks consisted of 5 serial span tasks based on different kinds of material:

1. Digit span
   Ex.: 2 / 8 / 6 / 5; 4 / 6 / 2 / 5 / 8 ; etc.
2. Pseudo word span:
   Ex.: vouleau / rindas / tuquin / prouline; coustère / possan / nibrol / miteau / nousette; etc.
3. Word span:
   Ex.: gratin / savon / soupape / tulipe; poulain / marteau / bûcher / ragoût / chaudière; etc.
4. Word span for words belonging to the same semantic categories
   Ex.: déluge / verglas / grêlon / tornade; canot / voilier / péniche / radeau / gondole; etc.
5. Word span for phonologically related words
   Ex.: gitan / ciment / milan / piment; colis / momie / rôti / croquis / taudis; etc.

Each of these tasks included 3 series of lists with 4 to 12 items. The lists were presented by series (cf. Desmette et al. 1995)—i.e. the subject first listened to a series of 4 items, then 5, then 6 and so on, until recall of all items in the right order failed. After failure the series was interrupted and the next series, starting with a 4-item list, was proceeded. The span for each task is the mean of the last lists correctly recalled for each series. It was preferred to base the study on means and not on the measures usually taken in the habitual bloc-presentation for memory span tasks, since means allow to take into account more subtle differences in performance.
The lists had been tape-recorded with a DAT (Digital Audio Tape) recorder by a single female native speaker with a very clear and unmarked pronunciation and were presented via professional quality headphones. Each list started with the signal '*attention*!' followed by the presentation of the list. The monosyllabic digits were presented at intervals of 1000 ms, the bi-syllabic words and pseudo words at intervals of 1250 ms[iii]. At the end of each list there was a discrete sound and the subjects had to recall immediately the words in the same order.
Written instructions were presented for each task followed by a training trial with one list of 4 items and one list of 5 items.

(b) Tasks involving storage and processing:

6. Free recall: subjects heard lists of 12 items (recorded in the same manner as for the serial span tasks) and had to recall as many of these words in the order they wanted. There were two conditions:
   a) after listening (3 lists plus one training list)
   b) after listening with articulatory suppression obtained by the repetition of the syllable 'bla' while listening (3 lists plus one training list).

7. A category and rhyme probe task: subjects heard lists of 4 to 12 items, similar to the word lists used in task 3, followed by a probe word presented 500 ms after the end of the list (indicated by the same sound signal). There were two conditions:
   a) *phonological condition*: tell whether the probe word rhymed with one of the words of the list; the response required was 'yes' or 'no'.

bifteck		*morue*	
panneau		*capote*	
tremplin		*glycine*	
boulon		*trousseau*	
canot	YES	*bourrique*	NO

   b) *semantic condition*: tell whether the probe word belonged to the same semantic category as one of the words of the list; the response required was 'yes' or 'no'.

jasmin		*colosse*	
reliure		*ballot*	
compresse		*fenouil*	
belette		*poussin*	
marmotte	YES	*chausson*	NO

   The position of the target word in the list was counterbalanced across lists. This led to the use of 8 series of stimuli, presented by series in the same manner as in the serial span tasks.

8. Listening span: A listening span task was recorded based on the French reading span version of Desmette et al. (1995). Participants were required to listen to sets of unrelated sentences, to repeat each sentence and to remember the last word of each sentence. At the end of each set, participants were asked to recall the last words in the same order. There were 3 series with sets ranging from 2 to 6 sentences. They were presented by series and rated by the means of the maximum performance for each series, in the same way as the other tasks.

(c) Stroop test:

9. A unilingual and bilingual Stroop[iv] task (Stroop 1935) in French and English: in the unilingual standard version of the Stroop (Golden 1978), participants have to read aloud, or name, three sheets as fast as possible within 45 seconds:
   a) Word score: Read aloud colour names printed in black ink
   b) Colour score: Name colour blocks
   c) Word/colour score: Name the colour of the ink of colour names which do not correspond to the colour of the ink (e.g. the word 'green' printed with red ink; the correct response being 'red').

In the bilingual version of the Stroop test (Preston 1965) the language in which the ink has to be named does not correspond to the language of the colour word. In all cases, scores correspond to the total number of words/colours named within 45 seconds. Only the word/colour score is taken into account here.

Procedure

Data collection took place between February 2002 and June 2004. Before testing, potential participants filled in a questionnaire investigating their professional experience, education, language skills and so on. This allowed to check for the eligibility of the participants with respect to the selection criteria (French A language, professional experience, etc.) and to gather additional data that might be of interest for subsequent analysis. Subjects were tested individually in at least two sessions in order to avoid fatigue. The whole procedure took around three hours per person. Task order was the same in all subjects since it was not clear in the beginning of the study if it would be possible to find enough participants for each group to allow for a real counterbalancing of order effects with a random presentation. Order of presentation in the two sessions was as follows:

Session 1:
Task 9: French Stroop test (word score, colour score and word/colour score)
Task 1: Digit span
Task 3: Word span
Task 4: Semantically related word span
Task 2: Pseudo word span
Task 5: Phonologically related word span
Task 6: Free recall
Task 9: French Stroop test (bilingual word/colour score)

Session 2:
Task 9: English Stroop test (word score, colour score and word/colour score)
Task 7a: Rhyme probe task
Task 8: Listening span
Task 7b: Category probe task
Task 9: English Stroop test (bilingual word/colour score)

The first session lasted about one and a quarter to one and a half hour, and the second around one and three quarter hour. Testing took place either in the subject's home or in a quiet room at their work place (or school).
The general hypothesis underlying the whole test design was that there would be evidence for a developmental evolution of working memory skills with interpreters performing better than both groups of control subjects, and interpreting students falling in between the two since they had just started developing these skills. It was expected that such differences would be more obvious in more complex tasks involving central executive capacities, such as tasks 6b, 7, or 8. Additionally we expected professionals to have a higher performance regarding selective attention as measured by the Stroop.

Results

For each task, the memory span measures were submitted to a one-factor analysis of variance with the between subjects factor 'Group' (Experts vs. Novices vs. Bilinguals vs. Control students).

These ANOVA did not reveal any significant differences between the groups in tasks involving storage only in the sense predicted by our hypotheses. The only task of type (a) yielding a significant group effect ($F=3.010$, $p=.0354$) was the word span (task 3). A Tukey test, however, shows that this effect arose from differences between the two control groups ($p<.05$) and did not concern the experimental groups we are interested in here (cf. appendix for more details).
The span tasks further replicated the tendencies usually observed in such tasks (cf. the data in appendix): digits were easier to recall than words, words were easier to recall than pseudo-words, semantic links between words slightly facilitated recall, and phonological similarity led to interference which brought about a lowering of performance (e.g. McCarthy & Warrington, 1994). There was rather little variation both within and across groups. All in all, these results confirm our expectations: the absence of significant differences between interpreters (novice and experts) and control subjects suggests that short term retention in the phonological loop does not play a major role in simultaneous interpreting.

In contrast, the results from the Stroop tests (c) do not confirm our hypotheses. The ANOVA did not reveal any significant differences between groups in the French and English unilingual Stroop tests (c) and in the English bilingual Stroop test (all data are given in appendix). The only group difference was obtained in the French bilingual Stroop ($F=4.560$, $p=.0146$) where the novice interpreters performed significantly better than both the expert interpreters and the bilingual control group as shown by the Tukey test ($p<.05$). This result, however, does not necessarily point to more highly-developed selective attention skills in novice interpreters. As the higher performance is obtained only in the French version of the bilingual Stroop (where the ink of English colour words is to be named in French) and not in the English version (where the ink of French colour words must be named in English) it is more likely to be indicative of a lack of balance between the languages in the novice interpreters, the dominant language being less affected by interference from the non-dominant language than the opposite.

In line with our hypotheses, significant group differences were found in the more complex tasks involving storage and processing, i.e. free recall with articulatory suppression (task 6b), the semantic probe task (task 7b) and the listening span.

Free recall (task 6)

Whereas in the control condition of this task (6a - free recall without articulatory suppression) the differences between the groups were not significant (cf. appendix), they reached significance ($F=2.819$; $p=.0447$) in the experimental condition where the subjects had to repeat the syllable 'bla' while listening to the word lists (6b - free recall with articulatory suppression) (see table 1).

Table 1. Mean number of words recalled per group in free recall with articulatory suppression (task 6b)

Group	Number	Mean	Standard	Standard error

			deviation	
Bilinguals	20	4.370	.805	.180
Students	20	4.415	.879	.197
Novices	18	5.244	1.402	.330
Experts	21	4.795	1.055	.230

The Tukey test revealed no further significant differences; however, the tendencies shown by the data suggest that the novice interpreters perform higher than the experts; the experts' performance falling in between the control groups' and the novices' means.

Rhyme and category probe task (task 7)

Contrary to the phonological condition of this task—which did not yield any significant differences between the groups (cf. appendix)—the group effect was highly significant in the semantic condition (F=8.372; p=.0001). Table 2 indicates that, again, the best results were obtained by the novices, but the three bilingual groups (bilingual controls, novices, and experts) performed significantly better than the control students (as shown by a Tukey test significant at p<.05). The differences between the three bilingual groups were not significant.

Table 2. Mean number of words recalled per group in the category probe task (task 7b)

Group	Number	Mean	Standard deviation	Standard error
Bilinguals	20	8.198	1.239	.277
Students	20	7.175	1.107	.248
Novices	18	9.200	1.304	.307
Experts	21	8.488	1.412	.308

Listening span (task 8)

The listening span task gave rise to a similar group effect (F=5.511; p=.0018). Again the novice interpreters had the highest scores, and the experts reached an intermediate level (table 3). The Tuckey test showed that statistical significance was due to the high performance of the novice interpreters which was significantly better than that of both control groups (p<.05). The difference between experts and novices was not significant, nor were those between experts and both control groups.

Table 3. Mean number of words recalled per group in the listening span (task 8)

Group	Number	Mean	Standard deviation	Standard error
Bilinguals	20	3.510	.816	.183
Students	20	3.445	.719	.161
Novices	18	4.539	1.205	.284
Experts	21	3.910	.927	.202

To summarize these findings, it is quite evident that short term retention based on the phonological loop as tested by serial memory span tasks is not more developed, neither in novice nor in expert interpreters. Similar findings were obtained with respect to selective attention as measured by the Stroop test. The more complex memory tasks used appeared to be clearer indicators for assessing the specificity of interpreters' cognitive skills. Significant group effects have been observed in free recall with articulatory suppression, the category probe task and the listening span, tasks that are more likely to invite the use of semantic strategies based in the central executive than the tasks which do not yield significant group effects. However, it is somewhat puzzling that the highest scores in these tasks were achieved by novice interpreters rather than experts (though the difference between the two groups of interpreters is never significant). These aspects will be discussed below.

Discussion

All in all these findings are not in contradiction with the other empirical studies discussed above.

As far as free recall with articulatory suppression is concerned, our findings confirm those from Padilla Benítez (1995) in spite of the important discrepancies regarding other aspects of both studies (e.g. the novice-expert distinction discussed below). These results seem to give some credit to the hypothesis that simultaneous interpreting between two oral languages (as opposed to sign language) generates phonological interference (e.g. Darò & Fabbro 1994; Isham & Lane 1994). The better performance under articulatory suppression would be indicative of the fact that the practise of the simultaneous interpreting task leads to greater resistance to phonological interference. Moreover, this finding suggests that simultaneous interpreters do not rely on phonological rehearsal since their memory performance is less disrupted by articulatory suppression. This is in line with the absence of any group effects in the serial span task, especially in those based on phonological rehearsal (pseudo-word span and word span for phonologically related words). It should furthermore be noted that our finding suggests that it is not exactly the degree of expertise in the interpreting task which leads to greater resistance to interference in the articulatory suppression condition, since the highest scores are not obtained by the expert group but by the novice group—contrary to the findings from Padilla. This aspect will be further discussed below.

With respect to the listening span, Liu et al. (2004) did not observe any difference between novice and expert interpreters. This is corroborated by our study: the group effect we found is due to differences between the novice interpreters and both control groups. Nordet & Voegtlin (1998) did not find any group effect in the listening span; however, the tendencies shown by their results are similar to our findings: interpreting students performed slightly better, professionals were intermediate and controls (students only in this study) had the poorest performance. It must be kept in mind, however, that this pilot study involved only a small number of participants which might be the reason for the absence of any significant result. On the other hand, our results are quite different from the reading span performance found in Padilla's 1995 study: in this study, the experts were significantly better than all other groups, including both novice groups. Padilla Benítez (1995: 126) mentions two possible reasons for the absence of any developmental effect in her data: firstly, she underlines that in Spain there is no pre-selection whatsoever before accessing the interpreting curriculum which starts earlier than in other countries; and secondly she claims that the acquisition of such cognitive skills may take more than one year of training. The

latter point is not corroborated by our results since the novice group in our study comprises students who were only in their second year of training but who had obtained the highest scores in most tasks. There are, however, important differences between both studies in relation with the type of participants involved. As stated above, Padilla's expert group is quite young and inexperienced and meets in these aspects our novice group rather than our experts. Since these two groups (Padilla's experts and our novices) actually behave similar, the difference in the findings could hence be attributed to participant characteristics.
The category probe task, which gives raise to a highly significant group effect, has not been used in other interpreting studies and so it cannot be compared to other findings. However, it should be noted that the performance patterns are slightly different in this task than in the other two: whereas the two control groups (students and bilinguals) behave in exactly the same way in both the free recall task and the listening span, the bilingual control group performs significantly better than the control students in the category probe task. This result suggests that the category probe task implies some specific skill which is developed not only in novice or expert simultaneous interpreters but also in highly skilled bilinguals.
Further analyses of the data should bring to light whether there are other factors which influence performance in these tasks, and, more specifically, the role played by age and different kinds of experience respectively. At the present state of data analysis, two aspects of the findings are of particular interest: a) the novice/expert distinction and b) the nature of the tasks where novice and—albeit to a lesser degree—expert interpreters score higher.

Novice/expert distinctions

The rationale underlying the experimental design was that simultaneous interpreting would rely heavily on specific cognitive skills which would develop with growing professional experience. Such a rationale is based on a conception of working memory close to capacity view (e.g. Just & Carpenter 1992) implying, schematically, that more competent individuals have more room to store information in working memory (cf. Hambrick 1998). Our findings strongly suggest that such a view does not hold with respect to simultaneous interpreting. Rather, our results sustain the idea that novice and expert processing are fundamentally different processes, as has been claimed by Moser (2000: 88) who summarizes Ivanova's (1999) findings: "*Qualitative differences exist in the processes that mediate expert interpreting, which does not appear to be constrained by the same cognitive limitations as those that apply to novice processing*." As has been suggested in order to explain the still indeterminate role of working memory in language comprehension (McCarthy & Warrington 1994: 345ff), short term working memory could play the role of an emergency resource intervening in cases such as the cognitive overload novices are confronted with when they start simultaneous interpreting. In other words, working memory capacity is more likely to develop in novices who are struggling with a new task (Cowan, personal communication). In interpreting experts, other types of processes may develop and replace working memory. However, it is clear that the novice and expert interpreters involved in this study do not simply differ with respect to their degree of expertise in simultaneous interpreting: not only are the novices younger than the experts (mean age 26,2 vs. 44,4 years), they might also have undergone a more discriminating selection procedure for admission than the experts 20 years earlier, possibly favouring high-span individuals from the beginning (cf. the remarks above concerning the differences with the participants involved in the Padilla study). Furthermore, the interpreting training in some schools involves different kinds of memory training including span tasks which could have influenced the performance of the novices in our test independently of their practise in simultaneous interpreting. Even motivation could play a role since the novices were very motivated for the tests as they were convinced that this type of exercise could help them to become good interpreters. Most of these factors are difficult or even impossible to establish *a posteriori* but should receive more attention in future studies. Some of these aspects are discussed elsewhere (Köpke & Nespoulous in preparation)

The nature of the task

The aim of this study was to assess the participants' performance in a broad variety of working memory tasks which allows the linking of performance to task characteristics. In this respect our findings clearly show that short term retention based on the phonological loop is not any better developed in interpreters (be it novices or experts). Conversely, it would appear that tasks involving the central executive are more likely to be in relation with some aspects of simultaneous interpreting—even though our results are not always significant, the tendencies clearly point in that direction.

However, one task that needs to be further commented on is the probe task (7a and 7b). Here we opposed a phonological condition (rhyme probe task) and a semantic condition (category probe task), which were both meant to involve the central executive. However, this seems to be the case only for the category probe task. Anecdotical evidence from the remarks of the participants during rhyme probe task reveals that most of them focus on the second syllable of the words and do not process the words as such. Hence their strategies are exclusively based on articulatory rehearsal in the phonological loop or might even benefit from attention-free sensory memory (Cowan 2000a) since the vowels can be supposed to play a major role in rhyme identification. This is completely different from what subjects do in the category probe task, where the stimuli words are processed semantically in all cases. Here subjects report either analysing the category of the stimuli while listening to the list, or visualizing the stimuli's referents and trying to link them together. Consequently, this task, clearly gives rise to a semantic processing of the stimuli words. Similar strategies were reported by the participants in the listening span task. This draws attention to another difference between tasks which do not give rise to group differences and those which do: not only are the latter based on more central aspects of working memory, they also involve deeper processing based on semantic aspects of the stimulus words. This is in line with evidence observed with other paradigms such as the verbal fluency task (Casado & Jiménez 1996; Moser-Mercer, 2000: 89; Moser-Mercer et al., 2000) showing that expert interpreters are advantaged in such a task only if they can rely on semantic cues (as opposed to phonological or morphological cues), at least when the semantic fields involved cover areas that are part of their professional experience. This would mean that the semantic characteristic of the task is essential, and further suggest that this type of cognitive skill is not transferable from one task to another. Hence, there appears to be a strong task dependency implying that even a high degree of expertise acquired in one specific task will not necessarily spread over to similar tasks.

Another point which needs to be discussed in this context are the findings from the Stroop-test. The motivation for choosing this test was based on the idea that attention could play several roles in a complex task like simultaneous interpreting (Cowan 2000a: 125): not only in switching attention from one task to the other or in concentrating attention on the least automatised aspects of these tasks, but also in the inhibition of irrelevant information. This last aspect is exactly what is measured in the Stroop. The absence of the expected group effect in our Stroop measures can be explained in at least two ways. First of all, as mentioned above, the broad rationale of this study is that the human mind is made up of a set of distinct capacities that can be selectively disrupted in pathological cases and particularly developed in experts of a specific domain. While we assume that this rationale is valid, the instruments used to measure pathological disturbances may not be appropriate for measuring expertise and this might be specifically the case for the Stroop test. Secondly it can be argued that the Stroop test involves attentional skills that are rather remote from those involved in simultaneous interpreting. The Stroop indeed relies on visual material (as opposed to oral language interpreting dealing exclusively with auditory material) which can only be considered as having a very reduced semantic content. Hence it is most likely that the Stroop test is not the best instrument for measuring the type of attentional skills involved in simultaneous interpreting.

Theoretical implications

One thing that can be ascertained in view of these data is that, obviously, the role of working memory in interpreting is not simply to store phonological forms until they can be processed (cf. the analysis of Hulme 2000). Our data clearly demonstrate the absence of any differences between interpreters (be it novices or experts) and the control groups in tasks that rely on articulatory rehearsal in the phonological loop. This is not surprising as the interpreting task involves simultaneous speaking and listening which should prevent resorting to articulatory rehearsal anyway. The three tasks which give rise to group effects certainly imply a central working memory holding area which could be the central executive (Baddeley 1986) or the focus of attention (Cowan 2000b, Cowan et al. 2005). Cowan's distinction between attention free and attention dependent storage could be promising for distinguishing between the tasks where they behave exactly like the controls and those where interpreters have higher scores (Cowan et al. 2005: 49). But the latter tasks also allow the participants to rely on deeper semantic processing which is not possible in the serial span tasks involving very short delays. Nevertheless, these aspects of working memory seem to be more developed in the novice interpreters than in the experts (even though there are no significant differences between these groups). In our opinion, this points towards a high specialisation of the skills involved in simultaneous interpreting, which are no longer dependent on working memory once a certain degree of expertise attained. Such expert skills could be subserved by long term memory via the episodic buffer (Baddeley 2001), by the means of specific routines (Ericsson & Kintsch 1995), or by highly specialised schemes (Sweller 2003). The relatively low performance of highly competent professional interpreters in many of the memory tasks we used corroborates Ericsson's (1998) claim (quoted by Moser-Mercer 1998: 89) that "*high-level skills of experts are not immediately transferable to other domains and that experts forced to perform in an unfamiliar environment are like fish out of water: they will revert to being novices*." The question remains, however, which experimental tasks might be good indicators of such expert skills. This study clearly shows that simple span tasks will not tell us anything more. It also draws attention to a big difficulty encountered in complex tasks involving both storage and processing: as our analysis of the probe tasks shows, such tasks may be resolved in several ways and most of the time, we do not know how the subject did it. One step further could be to concentrate in future studies on the scope of attention and take inspiration from some of the tasks reported on in Cowan et al. (2005).

Conclusion

The aim of this research was to open the discussion on cognitive aspects of simultaneous interpreting with a relatively broad survey involving a large variety of working memory measures submitted to a relatively large number of participants. The findings from this first study are not unequivocal. This should not be surprising in such a —still!—new field. And it is rather natural if we consider the complexity of the simultaneous interpreting task under investigation. A complex task, by definition, is likely to involve quite a number of subskills. The risk arising from that is that there might be more than one road leading to Rome, and there might be different ways of being a good interpreter. But this should not prevent trying to discover the different possible roads!

Notes

---

[i] All interpreting students attended schools where the interpreting curriculum extends over 2 years at the master level. The first year of training is mostly devoted to consecutive interpreting, training in simultaneous interpreting starts at the end of the first year or at the beginning of the second.

[ii] Frequency has been established with the Brulex database (Content, Monsty & Radeau 1990). All selected items have a frequency of 1 to 1200 tokens.

[iii] These regular time intervals were obtained by presenting the lists visually to the reader with the help of the Superlab experimental software (Cedrus Corporation 1991).

[iv] The Stroop test has been chosen despite its visual character because it is simple and powerful (Milham et al. 2002). Furthermore, working memory has been assessed visually in other studies with interpreters (e.g. Padilla 1995) and there seems to be no difference between auditive and visual measures of such skills (Daneman & Carpenter 1980). It would however be very interesting to complete the investigation with auditory attention measures in future studies.

[*] **Acknowledgements:** We thank the participants and the institutions that made testing possible, namely the Directorate General for Interpretation at the European Commission (SCIC, Bruxelles), and the following interpreting schools: Institut Supérieur de Traducteurs et Interprètes (ISTI, Haute Ecole de Bruxelles), Institut Libre Marie Haps (ILMH, Bruxelles), Ecole Supérieure d'Interprètes et de Traducteurs (ESIT, Paris). This research was supported by an ACI grant (LACO 47) from the French Research Ministry. We are grateful to Nelson Cowan for his very helpful comments and suggestions and to one anonymous reviewer for the remarks which have greatly improved an earlier version of this paper. We would further like to thank Nathalie Panissal, Christiane Soum, and Carol Serrurier for advice at different stages of this research.

References

Baddeley, A. (1986). *Working memory*. Oxford: OUP.

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

Baddeley, A. (2000b). Working memory and language processing. In B. Englund Dimitrova & K. Hyltenstam (eds.) *Language Processing and Simultaneous Interpreting. Interdisciplinary perspectives* (pp. 1-16). Amsterdam: John Benjamins.

Casado, B. & Jiminez, J. (1996). *Mental processes involved in conference interpreting: Verbal fluency*. Unpublished M.A. thesis, Geneva: Ecole de Traduction et d'Interprétation, Université de Genève.

Cedrus Corporation (1991). *SuperLab, Release 1.4*. Maryland: Cedrus Corporation.

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

Content, A., Monsty, P. & Radeau, M. (1990). *BRULEX : une base de données lexicales informatisée pour le français écrit et parlé*. Laboratoire de Psychologie Expérimentale, Université Libre de Bruxelles.

Cowan, N. (2000a). Processing limits of selective attention and working memory. Potential implications for interpreting. *Interpreting*, 5 (2), 117-146.

Cowan, N. (2000b). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24, 87-185.

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

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

Darò, V. & Fabbro, F. (1994). Verbal memory during simultaneous interpretation: Effects of phonological interference. Applied Linguistics, 15 (1), 365-381.

Desmette, D., Hupet, M., Schelstraete, M.-A. & Van der Linden, M. (1995). Adaptation en langue française du « Reading Span Test » de Daneman & Carpenter (1980). *L'Année Psychologique*, 95, 459-482.

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

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

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

Golden, C. J. (1978). *Stroop color and word test*. Chicago: Stoelting Company.

Hambrick ,D. Z., (1998). *Summary of Ericsson & Kintsch* (1995) [online] <http://www.cc.gatech.edu/jimmyd/summaries/ericsson1995.html> (accessed november 2004)

Hulme, C. (2000). Language processing mechanisms and immediate memory: Possible implications for simultaneous interpreting. *Ascona II Workshop on Complex cognitive processes: simultaneous interpreting as a research paradigm,* July 1-6-2000 Ascona, Switzerland. [online] http://mambo.ucsc.edu/ascona/announcement.html (accessed 15 July 2004)

Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D. A., Martin, S. & Stuart, G. (1997). Word-frequency effects on short-term memory tasks: Evidence for a reintegration process in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23 (5), 1217-1232.

Isham, W.P. & Lane, H. (1993). Simultaneous interpretation and the recall of source-language sentences. *Language and Cognitive Processes*, 8: 241-264.

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

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

Köpke, B. & Nespoulous, J-L. (in preparation). *Listening span and free recall measures in simultaneous interpreters: no evidence for higher cognitive skills*. (Working title).

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

Martin, N., Saffran, E.M. (1997). Language and auditory-verbal short-term memory impairments: Evidence for common underlying processes. *Cognitive Neuropsychology*, 14 (5), 641-682.

Martin, R.C., Shelton, J.R. & Yaffee, L.S. (1994). Language processing and working memory: Neuropsychological evidence for separate phonological and semantic capacities. *Journal of Memory and Language*, 33, 83-111.

McCarthy, R.A. & Warrington, E.K. (1994). *Neuropsychologie cognitive. Une introduction clinique*. Paris: Presses Universitaires de France.

Milham. M.P., Erickson, K.I., Banich, M.T., Kramer, A.F., Webb, A., Wszalek, T. & Cohen, N.J. (2002). Attentional control in the aging brain: Insights from an fMRI study on the Stroop task. *Brain and Cognition*, 49, 277-296.

Moser-Mercer, B. (2000). Simultaneous interpreting. Cognitive potential and limitations. *Interpreting*, 5 (2), 83-94.

Moser-Mercer, B., Frauenfelder, U.H., Casado, B. & Künzli, A. (2000). Searching to define expertise in interpreting. In B. Englund Dimitrova & K. Hyltenstam (eds.) *Language Processing and Simultaneous Interpreting*. Amsterdam: John Benjamins, 107-132.

Nordet, L. & Voegtlin, L. (1998). *Les tests d'aptitude pour l'interprétation de conférence. La mémoire*. Mémoire pour l'obtention du Diplôme de Traducteur, Université de Genève.

Padilla Benítez, P. (1995). *Procesos de memoria y atención en la interpretación de lenguas*. Tesis Doctoral, Facultad de Filosofía y Letras, Universidad de Granada.

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

Preston, M.S. (1965). *Inter-lingual interference in a bilingual version of the Stroop colour-word task*. Unpublished Ph.D. Dissertation, Montréal: McGill University.

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

Sweller, J. (2003). Evolution of human cognitive architecture. In B. H. Ross (Ed.) *The psychology of learning and motivation vol. 43* (pp. 215-266). New York: Academic Press.

Walker, I. & Hulme, C. (1999). Concrete words are easier to recall than abstract words: Evidence for a semantic contribution to short-term serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25 (5), 1256-1271.

Appendix

a) Simple span tasks

	Task	Bilinguals (N=20)	Students (N=20)	Novices (N=18)	Experts (N=21)	ANOVA 1
1	Digit span	5.98* .77	5.57 .74	6.18 .99	5.99 .74	F=1.91 p=.135
3	Word span	4.87 .63	4.39 .40	4.75 .57	4.70 .46	F=3.10 p=.035
4	Word span (same semantic category)	5.05 .50	5.00 .39	5.28 .46	5.09 .38	F=1.46 p=.232
5	Word span (phono-logically related)	4.16 .53	3.81 .81	4.30 .70	4.14 .51	F=2.02 p=.119

* Data indicate the mean number of words recalled in the right order and standard deviations.

	Task	Bilinguals (N=20)	Students (N=20)	Novices (N=18)	Experts (N=21)	ANOVA 1
2	Pseudo word span	1.50* 1.47	.95 .76	1.56 1.34	1.90 1.04	F=2.28 p=.086

* Data indicate mean scores and standard deviations. The pseudo-word task turned out to be very difficult for all participants, even the 4 item lists were in many cases not recalled which entailed an important floor effect. In order to avoid this, scores were not based on the number of words recalled, instead, subjects were scored one point for each list recalled correctly.

b)

	Task	Bilinguals (N=20)	Students (N=20)	Novices (N=18)	Experts (N=21)	ANOVA 1
6a	Free recall without art. suppression	5.33* .93	5.43 1.00	6.11 1.72	5.99 1.04	F=2.07 p=.111
7a	Phonological judgement	8.18 .97	7.64 1.15	8.34 1.38	7.94 1.33	F=1.22 p=,308

* Data indicate mean number of words recalled and standard deviations.

c) Stroop colour test

Task	Bilinguals (N=20)	Students (N=20)	Novices (N=18)	Experts (N=21)	ANOVA 1
Word/colour score French	49.55* 9.46	49.85 7.48	54.22 11.52	49.52 10.25	F=1.03 p=.385
Word/colour score English	50.70 8.07	n.a.	53.94 9.69	48.40 9.78	F=1.62 p=.207
Word/colour score bilingual French	60.15 9.39	n.a.	70.44 13.42	60.52 12.47	F=4.56 p=.015
Word/colour score bilingual English	55.15 8.53	n.a.	60.50 8.22	55.00 10.49	F=1.99 p=.146

* Data indicate the mean number of words (in 45 sec.) and standard deviations.