Simultaneous Interpretation & Working Memory

In simultaneous interpreting (SI), an auditory message in a source language is orally translated into a target language as the message continues. Rather than word-by-word translation, which is essentially impossible across different languages, there is a focus on providing the meaning of a message. The interpreter’s goal is to translate the important aspects of a message. The SI process taxes working memory (WM) in that the interpreter must concurrently keep the input in mind until a sufficient amount permits the interpreter to comprehend it, translate it, and produce a response at a quick rate. To perform this demanding task successfully, most professional simultaneous interpreters (PSIs) require years of training and practice. Some researchers maintain that PSIs have better WM as a result of their extensive training and practice in the task of SI (e.g., Darò, 1994). There is inconsistent evidence in the literature, however, as to whether or not PSIs actually have better working memory (WM) skills than non-interpreter multilinguals (NIMs).

This study attempts to understand why contradictory evidence has been reported by looking at the separate capacities within WM, as described by Randi Martin and her colleagues (e.g., Martin, Shelton and Yaffee, 1994), as independently from one another as possible. Martin and colleagues found a dissociation for phonological (speech sound) and semantic (meaning) that is well supported in the general literature. Most of the literature regarding WM in interpreters, however, has not addressed this dissociation. It is possible that memory differences between interpreters and non-interpreters may lie in these individual capacities of either phonological or semantic working memory. Therefore, if these capacities are not individuated cognitively, performance differences may have been camouflaged in previous studies.

Martin and colleagues (i.e., Martin, Shelton and Yaffee, 1994), profiled two brain-damaged patients with comparable decreases in WM span (i.e., the number of items that can be retained in WM). The patients were college educated and assumed to be English speakers. No other information was reported. One patient’s memory span, however, was relatively more impaired for phonological information than for semantic information whereas the other patient’s memory span was relatively more impaired for semantic that for phonological information. The work of
Nadine Martin and colleagues also supports the modularity of phonological and semantic representations in WM. Studying aphasic and normal language participants, Martin, Ayala, and Saffran (2002) and Martin and Saffran (1997) found evidence that semantic information influences primacy effects (i.e., better recall of list-initial items) and phonological information influences recency effects (i.e., better recall of list-final items).

The dissociability of WM is also seen with intact cognitive systems. Henk Haarmann and colleagues, for example, demonstrate the dissociation of semantic WM from other forms of working and long-term memory with their Conceptual Span Task (Haarmann, Davelaar, & Usher, 2003; Haarmann & Usher, 2001) and semantic category cued recall (Haarmann & Usher, 2001). Haarmann and Usher for example, found that word length effects (i.e., better recall for shorter relative to longer words) that are associated with phonological loop function were absent, in cued recall for semantic information. This suggests that semantic category cued recall invokes phonological rehearsal strategies to a lesser degree than an overt phonological tasks (e.g., serial order cued recall).

The modularity of WM is supported in neurophysiological studies as well. Crosson, Rao, Woodley, Rose, Bobholz, Mayer, Cunningham, Hammeke, Fuller, Binder, Cox, and Stein, (1999), using a carefully constructed control task, were able to locate different cortical areas for phonological versus semantic WM consistent with those cited in the literature via lesion studies. Areas of activation for semantic information included the left posterior inferior temporal lobe (near Brodmann’s areas 20 and 21) and the left inferior frontal gyrus, anterior to Broca’s area (near Brodmann’s area 47). Brain mechanisms for phonological information included the inferior temporal-occipital junction and an area of the prefrontal cortex (near Brodmann’s areas 44 and 46) independent of those seen for semantic activation. Studies with N-back tasks (i.e., where participants are presented with a series of items to remember and asked to recall one of the items depending on its position in the list) evidenced different areas of cortical activity for executive processing and phonological loop functions of WM. Smith and Jonides (1997) reported that activity for executive functions take place most dominantly in the dorsolateral prefrontal cortex, that rehearsal functions are localized to anteriorly to the premotor, that supplementary motor and Broca’s areas, and that storage is localized to post-parietal areas.

**Working Memory Models**

Among the models of WM best suited for addressing whether or not PSIs have better WM than NIMs is that of Baddeley and colleagues (e.g., Baddeley & Hitch, 1974; Baddeley 2000). Baddeley’s model is appropriate, in part, because it is consistent with the finding that there are
separate capacities for different types of information in working memory as mentioned above. Baddeley’s model also proposes specific components of WM that account for a variety of behavioral phenomena noted in WM tasks (e.g., recency, primacy, length affects). The Baddeley model, in particular, explains WM for lists of information, which has been the focus in the SI literature to date and will be the primary focus of this investigation. Other WM models, such as that of Just and Carpenter (1992), focus more on WM that involves higher-level language processing, such as sentence comprehension, rather than the simpler list-item memory.

The goal of this study is to understand WM differences between interpreters and non-interpreters on the most basic level of memory storage and processing despite the fact that interpreters work principally at a phrase or sentence level. Phrases and sentences require cognitive processes that include and exceed WM such as the processing of morphology (the grammatical units within a word), syntax (the grammatical relationships among words), and consequently, long-term memory. Sentences are common stimuli used for assessing WM skills. When sentences are used in a WM tasks, however, it becomes difficult to pin-point the exact cognitive process or processes where differences across varied groups manifest. Once basic WM processing performances in PSIs versus NIMs is clearly established, performance on higher level WM processing can be more appropriately addressed.

Baddeley and Hitch initially conceptualized a system involving three main components. This model included an attentional system known as the central executive and two subordinate systems. One system, the visuo-spatial scratch pad, processes visual and spatial information. The other system, the phonological loop, manages verbal information. The phonological loop consists of a phonological store where information is held for about 2 seconds before it begins to decay. The second component is a rehearsal process that refreshes information in the store via subvocal articulation. The central executive is a regulatory mechanism that controls information flow in WM, information retrieval from other memory systems, and information storage and processing in WM.

Recently, Baddeley (2000) has revised the model to include an episodic buffer. The buffer is a limited capacity storage system that is able to integrate various types of information such as visual and phonological information. This is important since many verbal WM tasks use visual stimuli. Similar to the components of the phonological loop, the episodic buffer is controlled by the central executive. The central executive can access the episodic buffer consciously and influence its contents by attending to an information source (e.g., LTM or the phonological loop). The episodic buffer was added to account for the recall and processing of prose and
semantic information as well. Baddeley proposes that the buffer is a mechanism for recalling the gist of a message rather than verbatim information.

In summary, there is ample evidence of the modularity of phonological and semantic information in working memory from the brain damage, cognitive psychological and neuropsychological literatures. There is also a distinction of phonological and semantic information in the SI process whereby semantic information may be more important to the SI process as it is the transfer of semantic information that is the ultimate goal of the process. Using the revised Baddeley model as a theoretical basis, we can look at semantic and phonological memory relatively independently. This should tell us how each contributes to the process of simultaneous interpretation.

**WM in PSIs: Evidence from the literature**

Many of the studies that have looked a WM in interpreters have used measures of memory span that reflect WM capacity. There are two basic types of span measures, those that use words and those that use sentences. In word-span tasks, WM capacity is established as the maximum number of items an individual is able to repeat back immediately following presentation of a list of words. In sentence-span tasks, also referred to as phrase- or reading-span tasks, by contrast, capacity is determined by the maximum number of sentences an individual can repeat the final word from following presentation. Word-Span measures are typically considered to be representative of phonological loop function because they show the affects of the phonological loop such as better recall for information at the beginning of a list (primacy), better recall for information at the end of a list (recency), and better recall for shorter relative to longer words (length effects). The phonological loop, you will recall, functions to store and refresh verbatim speech sound information in WM. Sentence span tasks generally reflect complex storage-and-processing in WM. Using sentence stimuli, then, would prevent one from determining precisely what may be contributing to group differences in basic WM (e.g., phonological storage, semantic storage, etc). Thus, use of word list stimuli, rather than sentences, permits for relatively isolated testing of WM function. That is, word stimuli, unlike sentence stimuli, allows for assessment of phonological WM relatively independent of semantic WM and assessment of general WM in a relatively independent of long-term memory.

**Evidence of Superior PSI Performance over NIMs**

Word List Stimuli

There is evidence in the WM literature that employs word lists that suggests that PSIs have better phonological loop function than do NIMs. Darò and Fabbro (1994) and Padilla, Bajo, Cañas,
and Padilla (1995), for example, report higher than average digit spans for interpreters relative to
non-interpreters. Darò and Fabbro studied a group of 24 student interpreters as they participated
in a series of digit span tasks. The student interpreters ranged in age between 22 and 27 (average
age 24.5) and had been in SI training between 12 and 60 months (average 29 months) for 10 to
40 hours a week. They were described as highly proficient in their native Italian as well as two
other languages among a set of four: German, English, Spanish, and French. The authors divided
the participants into two groups according to what language they worked in. Group A performed
the task in Italian and Group B performed in English. No details were reported regarding how
group placement was determined.

Darò and Fabbro (1994) presented strings composed of digits from 1-9 binaurally over headsets.
The series began with two sets of three digits, then two sets of four digits, up to two sets of 9
digits. Participants immediately recalled the numbers in the order they had been presented after
the end of each set. The task yielded a span of 8.38 for Group A and a span of 8.90 for Group B.
The combined average span was 8.65. These figures are higher than reported average of 7 digits
in non-interpreters (Darò, 1994) and support the proposal that PSIs have better phonological loop
function relative to non-interpreters.

In another study that compared PSIs, student interpreters, and NIMs, Padilla et al. (1995) found
that PSIs had superior digit spans. They studied four groups of bilingual participants with varied
experience with SI. One group of ten participants were practicing interpreters, five who were
recent graduates of the School of Translators and Interpreters at the University of Granada and
five who had been practicing in the field for about five years. These individuals had a mean age
of 30 year. No other age related information was reported. A second group of ten participants
were third-year interpreting students described simply as having some SI training. No age
information was reported for this group. A third group of ten participants were also students.
This group of students, however, was in their second year of their program, and had training in
translation but not SI. Again, no age information was reported for this second group of students.
A final, fourth group of ten participants were bilinguals with no translation or interpreting
experience. Like the student groups, no age information was reported. There was also no other
information regarding background history of the participants was provided.

As in the study of Darò and Fabbro (1994), Padilla et al.’s (1995) participants listened to sets of
digits and were asked to recall them in order of presentation. The task commenced with three
sequences of four digits. Each set of sequences increased by one digit so that the second set had
three sequences of five digits. Sets were presented in these increasing sizes until a participant
was unable to recall the digits in order. The results indicated that the PSI group had significantly longer digit spans than the other groups. Although the group digit-span means were not directly reported, graph data indicate that the PSI group had spans just over 7 digits where the other groups had spans just over 6 digits. The authors note that the second-year students, without SI training, performed comparably to the NIM group.

Another recent study (Christoffels 2004) looking at interpreters and highly proficient bilinguals also found statistically superior PSI performance using word stimuli as opposed to the digits used in the previously mentioned studies. The control group consisted of balanced Dutch/English bilingual university students who spoke Dutch as a first language. All had at least six years of formal education in English as a second language and used English textbooks for their university course-work. The experimental group of interpreters was also native Dutch speakers and used English as one of their languages of interpretation. No information was reported on the age of the participants.

Participants saw words, one at a time, on a computer screen, in sets of three to 10 words and were asked to recall them, in order, after the presentation of the final word in a list. The task was executed in both Dutch and English. The results revealed statistically superior performance by the interpreters over the non-interpreters with the former group achieving word spans of 5.0 and 5.92 in Dutch and English respectively and the latter group achieving word spans of 3.59 and 3.05 in Dutch and English respectively.

It is also worth noting superior PSI performance in word-list recall, a task that is somewhat similar to span measures like those mentioned above (i.e., word lists are presented and information from the list is recalled). Superior WM and phonological loop function in particular, was supported by Padilla et al.’s (1995) study in another set of experiments that employed list recall with articulation suppression. Articulation suppression is a method commonly employed to assess phonological loop function. This method assesses the storage component of the phonological loop because it prevents subvocal articulation in the phonological loop to refresh information in WM.

The four Padilla et al. experimental groups participated in a series of word-list recall tasks with articulation suppression. In these tasks, participants read three lists of 16 printed words, presented at the rate of one word every three seconds. They articulated the nonsense syllable, “bla” repeatedly during list presentation. In a free-recall condition, participants recalled each
word list at the conclusion of list presentation. In a final-recall task, participants recalled as many words as possible after all three lists had been presented.

The PSI group recalled significantly more words than the non-PSI groups in both free and final articulation suppression recall tasks. The raw data were not reported, but graph data indicate that the PSI group recalled approximately 55 words relative to the other groups whose performances approached only 40 words. This suggests that PSIs have better WM skills than non-PSIs, at least when the rehearsal process of the phonological loop is effectively removed from WM function. Padilla et al. (1995) ran this task without articulation suppression as well. The results of this second condition are discussed in the following section on performance parity.

Köpke & Nespoulous (2005) found a similar result under articulation suppression comparing professional simultaneous interpreters, student interpreters, bilingual speakers, and monolingual speakers. They used aural stimuli, however, in contrast to Padilla et al.’s (1995) print stimuli. Köpke & Nespoulous’ professional interpreter group consisted of 21 individuals who ranged in age from 29 to 61 years (mean age 44.4 years) and had four to 35 years of interpreting experience. The student interpreter group consisted of 18 individuals, ages 23 to 38 years (mean age 26.2), and who were in their second year of studies and had just begun their training in simultaneous interpreting. There was one control group that consisted of 20 bi- and multilingual speakers who ranged in age from 27 to 63 years (mean age 44.7 years) and a second control group of 20 monolingual speakers who ranged in age from 18 to 26 (mean age 21.5). All testing was done in French, the dominant language of all the participants. All the multilingual participants, save the monolingual group, were highly proficient in at least English and sometimes other languages. No other specifics were given to what the languages were, how they were used, how they were acquired, etc.

In the Köpke & Nespoulous (2005) free recall task, the participants heard word lists of 12 items while repeating the word “bla”. At the end of a list they recalled the words in any order. Surprisingly, the student interpreters had the best performance (mean of 5.2), followed by the professional interpreters (mean of 4.7), the monolingual speakers (mean of 4.37), and then the multilingual speakers (4.4). The authors report that group differences were significantly different, but do not specify further. Köpke & Nespoulous ran their free recall task without articulation suppression in addition to a number of other tasks. These will also be discussed in the following section on performance parity.
Another word list task, as mentioned earlier, is cued recall. Köpke & Nespoulous (2005) ran their participant groups in a cued recall task that revealed a significant, and interesting, group difference that suggests a potential advantage for interpreters over non-interpreters. In this task volunteers heard lists of four to 12 items long followed by a probe word. After listening to the word list, the participants responded, “yes” or “no” to indicate if the probe word belonged to the same semantic category as one of the words in the list. The novice/student interpreters performed significantly better than the other groups. The professional interpreters performed next best followed by the multilingual controls and then the monolingual controls. No differences were found between the professionals and the multilingual controls. There was a pattern of better performance, however, for the interpreters.

Sentence Stimuli
As mentioned above, sentences are used in span tasks to assess WM capacity in addition to word lists. Padilla et al. (1995) also ran their four groups of participants in a reading-span task. The task was a Spanish adaptation of Daneman and Carpenter’s reading-span task (1980). Recall that reading span is a complex-storage-and-processing task in contrast to the simple storage task aforementioned. Participants read sets of sentences presented one at a time then recalled as many final words from as many sentences as possible from a set. The number of sentences per set increases as the task progressed. The maximum number of final words recalled indicates what the authors called the participants’ phrase span. Padilla et al.’s PSI group had significantly longer phrase spans than the other groups. Graph data for phrase span indicated spans of 5 for the PSI group and spans of just under 4 for the other groups. Again, as with the digit span task, the second-year students without SI training from this study, performed comparably to the NIM group.

Christoffels (2004) also found superior PSI performance in a reading span task. Her participants read aloud 42 English sentences in sets of two, three, four, and five. After reading the final sentence in a set, a beep sounded and the participant would recall the last word of each sentence. Significant group differences manifested with the interpreters outperforming the non-interpreters with reading spans scores of 34.54 and 31.13 (out of a possible 42) respectively. Results from this task are potentially misleading, however, as a comparable version in Dutch revealed a different result. This latter outcome will be addressed in the following section.

A potential interpreter advantage was also found in a listening span task in the Köpke & Nespoulous (2005) study. The professional interpreters, student interpreters, multilinguals and monolinguals in this study listened to spoken sets of sentences. They repeated each sentence and
were asked to remember the last word of each sentence. After a set was presented, the participants recalled the last word from each sentence in the order of presentation. As with their semantic cued recall task, Köpke & Nespoulous found that the student interpreters out-performed the other three groups. This difference was significant for the student interpreters versus the control groups, but not for the students versus the professional interpreters or the professionals versus the controls.

In summary, there are both simple storage and more complex storage-and-processing tasks that show PSIs have better working memory skills relative to NIMs. Simple storage advantages for PSIs are further substantiated by tasks employing articulation suppression that prohibit phonological rehearsal to refresh information stored in WM. Advantages for PSIs relative to NIMs was also seen for stimuli that were both aural and print in nature.

**Evidence of Parity of WM Skill in PSIs & NIMs**

The studies and tasks reviewed above suggest that PSIs have better WM skills than NIMs. One must note, however, that there are other studies and tasks that suggest there may be no WM differences between the groups. While Darò and Fabbro (1994) found statistically significant group differences, Padilla et al. (1995), Christoffels (2004), and Köpke & Nespoulous (2005) found mixed results. Chincotta and Underwood (1998) and Liu (2001) found no significant differences all using similar tasks.

**Word List Stimuli**

As mentioned in the last section, Padilla et al. (1995) ran their four participant groups in free- and final-recall word-list tasks. When forced to speak during stimulus presentation, PSIs outperformed NIMs in their ability to recall the words presented. When the tasks were run in an alternate condition, however, wherein participants simply listened to the lists and then recalled them, either immediately (free-recall) or after a delay (final-recall) as described earlier, no group differences in the number of words recalled manifested.

Results of digit span tasks also appear to demonstrate parity of performance. Contrary to Padilla et al.’s (1995) digit span results of significant group differences using auditory stimuli mentioned in the previous section, Chincotta and Underwood’s (1998) findings indicate no significant differences in memory span for printed digits between student interpreters and NIMs. Chincotta and Underwood compared 12 student interpreters from the University of Turku with a minimum of 100 hours of experience interpreting between Finnish and English to a group of twelve
Finnish English Majors. The participants were all described to be Finnish dominant, Finnish-English bilinguals. No other information on language history or age was reported.

The participants saw four sets of digits presented one at a time, sequentially on a monitor. Each set began with two series of two digits and then two series of three digits increasing by one digit every series up to two series of 13 digits. Following a prompt tone at the end of each series, the participants orally recalled as many numbers as possible in both Finnish and English conditions. The stimuli were also presented in normal and suppressed conditions. In the suppressed condition, participants articulated the nonsense syllable “la-la” continually beginning four seconds prior to the initial stimulus presentation and ending four seconds after presentation of the last digit stimulus. The results indicated no significant differences between groups in mean digit span in either language.

Köpke & Nespoulous (2005) ran a series of word list tasks that, unlike most other studies looking at interpreters, addressed the dissociation of phonological and semantic WM. As mentioned earlier, they compared professional simultaneous interpreters, student interpreters, bilingual speakers, and monolingual speakers. Participants performed auditory word span tasks with real words, non-words, semantically similar words and phonologically related words.

In the span tasks, participants listened to word lists four to 12 words in length and repeated them back in order as best they could. Results of the span tasks revealed no significant differences across the groups. Also, as mentioned, they had performed a free recall task without articulation suppression. Recall that under articulation suppression group differences were found. Without articulation suppression, however, Köpke & Nespoulous, like Padilla et al. (1995), did not find significant group differences. Mean scores and statistics were not reported for the free recall without suppression.

Köpke & Nespoulous’s (2005) participants also performed phonological cued recall tasks to compare to the semantic cued recall task. In the phonological condition, after listening to the word list, the participants were given a probe word and had to report if the probe word rhymed with one of the words in the list by responding, “yes” or “no”. Contrary to the semantic condition where group differences were noted, the authors found no differences across their participant groups in the phonological task.

Sentence Stimuli
In addition to tasks using word-list stimuli, performance in tasks using sentences also suggests a parity of performance between interpreters and non-interpreters. Liu (2001) conducted a study that compared 13 professional interpreters, who had at least one year of training and two years of professional experience, to two groups of student interpreters. One group of 12 students had a year and a half of SI training. The other group of 11 students had just begun SI study. All the participants were native Mandarin speakers who spoke English as a second language. Some participants were described as having native-like proficiency in both languages. There was no indication of which participants this description pertained to, however. Liu noted that the students were enrolled in programs in Taiwan and California, though the number of students in each program was not reported. The working situation of the PSIs was also not included. No other details pertaining to language history or age were provided.

Liu’s phrase-span task is another adaptation of Daneman and Carpenter’s (1980) reading-span task. Instead of reading, her participants listened to a series of unrelated spoken English sentences 13 to 16 words in length. Each series consisted of five sets of sentences, beginning with two sentences and progressing to five sentences by an increase of one sentence per series over the course of testing. At the end of each set, the participants wrote down as many of the last words from the sentences as they could remember. The number of final words recalled indicated a participant’s phrase span. Liu found no significant differences in phrase span across groups. There was a pattern, nonetheless, for professionals to perform better than students and for students with more interpreting experience to perform better than those with less experience. Mean scores for professional interpreters, advanced students, and beginning students were 3.3, 2.91, and 2.59 respectively.

Sentence-span tasks using printed stimuli also support performance parity. As mentioned earlier, Christoffels (2004) ran their participants on a reading-span task in their native Dutch in addition to the English reported in the previous section. Despite significant group differences on the English condition of this task, no significant group differences manifested for Dutch. The mean reading span scores for the interpreters and the non-interpreters were 35.39 and 34.0 respectively out of a possible 42. It is worth noting, however, that like the other studies that suggest parity between interpreters and non-interpreters, the interpreter group’s performance in the Christoffels study is in the direction of PSI superiority as well. An important finding to highlight from this study is the performance difference in a first (L1) relative to a second language (L2). The interpreters in this task were at a significant advantage only when performing in their L2.
In sum, one set of studies evidences statistically significant superiority of PSIs over SI students and NIMs, with a subset showing student interpreters having the advantage over the other two groups, while a second set does not. The second set, however, often shows absolute differences in the direction of interpreter (student or professional) superiority in WM. Thus, one may conclude that the ‘no-difference’ findings may reflect a small but real effect in conjunction with the relatively small sample sizes and the range of inter-individual differences among participants. Moreover, that there were studies whereby the same participants either evidenced group differences or not, depending on the nature of the stimuli and the task (e.g., Padilla et al., 1995; Köpke & Nespoulous’, 2005) suggests that task differences are crucial in determining whether group differences will be manifest.

Why the Contradictory Findings?
There are a number of possibilities as to why some studies indicate better WM function in PSIs in comparison to NIMs and others do not. These may include participant selection (e.g., the number of participants, the languages they spoke, amount of SI training, proficiency, age at testing, and use of languages tested), the nature of the stimuli (e.g., digits versus words, printed versus aural presentation, and cross-linguistic issues such as longer versus shorter articulation duration of stimuli), and the procedures (e.g., tasks to assess the various components of WM such as the phonological loop versus the episodic buffer).

Consider, first, the participants selected to take part in each experiment. Many of the studies surveyed here had groups as small groups. A small number of participants does not yield strong statistical power and compromises external validity. This presence of a small N is important to note since individual performances varied and there was a frequent pattern of superior performance with increased levels of SI experience relative to less or no SI experience across the reported studies. It stands to reason that these patterns might have yielded statistically significant differences had the groups been larger.

Consider, also, the limited background information and perhaps control regarding language history as a contributing factor to the contrastive findings across studies. Pertinent information to collect and control for would include the manner in which languages were acquired, the order of language acquisition, the nature and degree of language use, and the phonological, orthographic and morpho-syntactic nature of the languages that the participants know. Insufficient control on these variables contributes to heterogeneity of the participant groups and exacerbates the problems of a small N. Collecting such information is also important because these variables can influence language proficiency. Ascertaining proficiency is important because competency in a
particular language influences WM performance in that language Thorn and Gathercole (2001), for example, show that WM capacity measured by nonword repetition is related to proficiency as measured by vocabulary knowledge in the language of testing. The importance of the proficiency issue regarding interpreters has also already been demonstrated. You will recall that the findings in the Christoffels’ (2004) study revealed significant differences between PSIs and NIMs were found in sentence span tasks when performing in L2 and not L1, despite high L2 proficiency across both groups.

Another aspect regarding background that may account for the discrepancy in findings is the age range of the participants. The literature shows that, as we age, certain aspects of language performance, like lexical retrieval and semantic short-term memory, deteriorate (e.g., Nicholas, Obler, Albert, & Goodglass, 1985; Haarmann, Ashling, Davelaar, & Usher, 2005). The age backgrounds of the participants in most of the studies surveyed above were not reported adequately to make proper speculations regarding age as a performance factor. Since age can influence performance it should be controlled for carefully in cognitive and linguistic experiments.

Cross-linguistic differences are also noted to manifest in WM tasks. Memory span, for example, is sensitive to word length. Welsh-dominant bilinguals evidenced greater digit span in English than in Welsh because Welsh digit names are longer than English digit names (Ellis & Hennelly 1980). We also saw cross-linguistic differences in the Darò and Fabbro (1994) study mentioned earlier, where Group A, who performed the task in Italian, had a slightly lower span than Group B, who performed the task in English. Italian digits have more syllables than their English counterparts so it is not surprising that Darò and Fabbro’s results are similar to those of Ellis & Hennelly.

The nature of the stimuli and how they were employed may also account for the inconsistent findings. Recall, there are no consistent patterns that emerged as the result of the use of a particular stimulus form (i.e., written versus spoken) or stimulus type (e.g., words versus digits versus phrases). Printed stimuli, at times, yielded no significant group differences between PSIs and NIMs. This is seen in the Chincotta and Underwood (1998) study with printed digits in both normal and AS conditions and in Padilla et al. (1995) with printed words in normal conditions. At other times, printed stimuli did reveal group differences where PSI did significantly better than non-PSI groups. This is seen in Padilla et al. with printed words under AS and with printed phrases under normal, non-AS conditions.
Spoken stimuli showed a similar inconsistency to that of print stimuli. Spoken digits from the Padilla et al. (1995) study revealed significant group differences wherein PSIs performed significantly better than NIMs. Spoken words resulted in mixed findings for Köpke & Nespoulous’ (2005). Spoken phrases were also mixed for Köpke & Nespoulous’. Liu (2001) found, however, that spoken phrases did not yield significant group differences. Liu found only a pattern for PSIs to perform more successfully than NIMs that did not reach statistical significance.

The nature of word list stimuli in regard to their semantic characteristics, and consequential semantic or other cognitive processing outside WM, should also be considered when attempting to account for varied findings in the literature. This is particularly important, for example, for the contradiction within the Padilla et al. (1995) study (i.e., significantly longer spans for PSIs than NIMs with words but not digits). Stronger semantic relations among word stimuli that belong to a large open set of options, relative to digits that come from a small closed set of options, may have also contributed to the disparity.

The word stimuli would have a greater number of semantic associations to facilitate recall relative to digits and non-words. The literature shows that in incidental or delayed recall, such as free-recall used in the Padilla et al. (1995) study, information is better recalled where salient semantic connections can be made relative to where few to no overt connections exist (e.g., Kroll & Stewart, 1994). This may explain, for example, why no group differences manifested with the word stimuli in the Padilla et al. study. That is, recall for the NIM groups under normal conditions may have been facilitated by semantic connections that, in turn, resulted in a performance on a par with that of the PSIs in the normal condition. The contribution from the semantic associations and consequential long-term memory support may have served a compensatory function and camouflaged a relative weakness in WM for the NIMs compared to the PSI group, should such a WM weakness exist. This support from semantic information may not have persisted in the suppression condition where we saw a stronger WM performance in the PSI group relative to that in the NIMs.

The nature of the stimuli in terms of being auditory or visual may account for the performance parity among the participant groups in the Köpke & Nespoulous’ (2005). Their attempts to separate phonological from semantic WM may have been undermined by using auditory stimuli. Auditory stimuli will automatically engage the phonological loop making semantic tasks have a strong phonological nature. Use of print stimuli may be a way of minimizing phonological
processing in a semantic task and a means of equating tasks so that the best picture of isolated phonological versus semantic WM can be evaluated.

Procedural differences may also account for the discrepant findings across groups. The free- and final-recall tasks of the Padilla et al. (1995) study, because of the delayed nature of recall, do not depend as heavily on phonological information or WM as the immediate repetition involved in span tasks. This procedural difference resulting in the activation of other non-WM processes, therefore, may account for the discrepancy within the Padilla et al. study whereby a significant group difference was obtained in immediate unsuppressed recall of digits but not in delayed unsuppressed recall of words.

In summary, a variety of factors including small group sizes, group heterogeneity, the varied nature of stimuli, the varied nature of stimulus presentation across studies, and the varied ages of the participants may account for the discrepancy regarding whether or not PSIs have better WM skills than NIMs. It would be prudent to consider and control for these factors in future research when attempting to determine if there are differences in WM between interpreters and non-interpreters and the nature of such differences should they exist.

References


Acknowledgements
Special thanks are due to Loraine K. Obler and Martin R. Gitterman for their input on the content and form of this paper throughout its development.