DECREASES IN WORKING MEMORY CAPACITY FOR SENTENCE STIMULI WITH ADULT AGING

A Thesis
presented to
the Faculty of the Graduate School
at the University of Missouri-Columbia

In Partial Fulfillment
of the Requirements for the Degree

Master of Arts

by
AMANDA L. GILCHRIST

Drs. Nelson Cowan and Moshe Naveh-Benjamin,
Thesis Supervisors

DECEMBER 2007

The undersigned, appointed by the dean of the Graduate School, have examined the thesis entitled

DECREASES IN WORKING MEMORY CAPACITY

FOR SENTENCE STIMULI WITH ADULT AGING

presented by Amanda L. Gilchrist,

a candidate for the degree of master of arts,

and hereby certify that, in their opinion, it is worthy of acceptance.

Professor Nelson Cowan

Professor Moshe Naveh-Benjamin

Professor Bruce Bartholow

Professor Linda Day

ACKNOWLEDGEMENTS

Many thanks to Drs. Nelson Cowan and Moshe Naveh-Benjamin for their assistance in forming this study, as well as providing their input during the various stages of statistical analysis and writing. Acknowledgements are also extended to Dr. J. Scott Saults, in helping to program the study; Drs. Bruce Bartholow and Linda Day, for their helpful suggestions for interpreting current findings, as well as forming future studies; and David Chen, Melissa Knipe, Alice Andres, Allison Federici, and Geoffrey Maddox, for their help in running participants.

This work was supported by a grant from the Roger S. Williams fund.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Previous studies have found that older adults have poorer immediate recall for language. Older adults may recall fewer chunks from working memory, or may have difficulty binding words or sentences together to form multi-unit chunks. We examined these two hypotheses by presenting participants with 4 types of spoken sentences for immediate free recall, varied by the number and length of chunks per trial: 4 short, simple sentences; 8 such sentences; 4 compound sentences, meaningfully comprised of two short sentences; and 4 random sentences. On those trials where the number or length of presented chunks were large, older adults formed fewer chunks than young adults with few differences in chunk size. Our results suggest that both of the above hypotheses influenced recall for older adults, though the degree of involvement of each has yet to be determined.

A fundamental hypothesis in cognitive psychology is that the working memory system is capacity-limited in terms of the number of chunks that can be stored or maintained at once (Miller, 1956; Cowan, 2001). The sizes of chunks can be estimated, allowing an estimate of the number of chunks held in working memory (e.g., Cowan, Chen, & Rouder, 2004). It is possible to ask, then, whether the number and size of chunks change with age in adulthood. Naveh-Benjamin, Cowan, Kilb, and Chen (2007) recently provided evidence from the serial recall of word lists that the number of chunks that can be held in working memory decreases dramatically in adult aging, and that the ability to use multi-word chunks also decreases somewhat. The present research investigates similar processes in natural language, making use of the idea that the coherence of language does result in multi-word chunks in memory (cf. Tulving & Patkau, 1962).

To provide insight into the motivations for the current study, a review of relevant literature begins with the phenomena of chunking and its relation to capacity limits in working memory. This will be followed by a discussion of aging and its effects on memory, with an emphasis placed upon how such effects might relate to chunk formation in language processing.

Working Memory and Chunking

The most notable conception of capacity limits in immediate memory is "the magical number seven plus or minus two (Miller, 1956)." Research has provided further evidence for a basic capacity limit in working memory, though somewhat smaller than originally conceived, reaching a limit of approximately three to four items, recalled without error or pause (Broadbent, 1975; Cowan, 2001). Cowan (2001) proposed that this was due to limits in the number of separate items that can be held in the focus of attention when precautions are taken to ensure that these items are not rehearsed or grouped together to form higher-level chunks. The focus of attention contains information in memory that is not only currently activated, but is also present in a person's current awareness or level of attention during completion of a task (Cowan, 1999, 2001, 2005). If a person is able to devote all attention to storing objects into memory, as opposed to performing a secondary task simultaneously, a greater number of items to be remembered can be active in the focus of attention.

Miller (1956) was the first to emphasize a feature of immediate memory that has had implications for working memory research. More items could be recalled when they were organized into superordinate groups on the basis of pre-existing knowledge or associations, known as chunking. Miller did not specifically note that the limit of about 7 items could reflect a smaller number of chunks if these chunks can be rapidly formed during the presentation of a list, and this may account for the different estimate of immediate memory capacity observed by Miller versus Broadbent (1975) and Cowan (2001).

In an early study of chunking in verbal free recall, participants were presented with lists of words that approximated English text to varying degrees (Tulving & Patkau, 1962). Participants could recall the words in the list in any order, but the authors defined a chunk as a group of items that were recalled in the correct serial order. A limit in the number of chunks recalled was relatively consistent across all manipulations of approximation to English, ranging from 4 to 6; this provides further evidence for some basic capacity limit in working memory. It also appears that when associations between stimuli become increasingly meaningful, the size of a chunk increases. The more closely the presented words approximated English text, the larger the chunk became. While the chunk size varied across conditions, the number of chunks recalled remained roughly the same, regardless of degree of approximation of the stimuli to English text. Similar results have been found using different units of speech (e.g., words, sentences, proverbs) (Glanzer & Razel, 1974; Simon, 1974) and also with repeated presentations of word pairs (Cowan et al., 2004). Nevertheless, it still remains uncertain where the boundaries between two respective chunks lie and, consequently, the number of chunks that are actually being retrieved by a participant.

Under conditions of free recall, chunk limits tend to operate (Chen & Cowan, 2005). In contrast, effects of list length may be more likely to operate in serial recall of relatively short lists. This implies that chunk and length capacity limits are not mutually exclusive in working memory operations. Different limits for recall operate under specific situational circumstances, but it is chunk limits that are of particular interest for the present study, as this would provide further evidence for a capacity-limited working memory system. As Chen and Cowan suggested, capacity limits can be observed for lists that are too long to allow covert verbal rehearsal (i.e., lists that take longer than about 2 s to pronounce; see Baddeley, 1986), and long sentences to be featured in this study tend to exceed this limit.

Aging and Memory

Generally, older adults do not perform as well as younger adults on memory tasks, especially those that involve episodic memory. One particular memory deficit of interest is known as a binding deficit. It has been suggested that while recall for a particular feature parallels recall performance in young adults, older adults may have deficits in binding of multiple features of objects during short-term encoding, which is often critical for good recall (Mitchell, Johnson, Raye, Mather, & D'Esposito, 2000; Cowan, Naveh-Benjamin, Kilb & Saults, 2007) These deficits are also found in episodic memory when two objects (e.g., word pairs, pictures) must be associated into a single unit (Naveh-Benjamin, 2000; Naveh-Benjamin, Hussain, Guez, & Bar-On, 2003; Naveh-Benjamin, Guez, Kilb, & Reedy, 2004; Naveh-Benjamin, Guez, & Shulman, 2004). Naveh-Benjamin (2000) proposed the associative deficit hypothesis (ADH) to account for the above findings, where long-term memory for component objects or features is relatively unaffected by aging, but the ability to form or memorize associations between features or objects significantly declines.

A chunk can be conceived of as a special type of binding or association, in which two or more pieces of information are bound together, perhaps with an additional binding to the current temporal context. According to the ADH, it would be expected that older adults will not do as well as young adults in forming cohesive chunks. This may lead to smaller chunks being formed, fewer chunks being formed, or both possibilities.

Beyond binding and associative deficits, there are other posited deficits or explanations for impairment in aging literature that may have a connection with chunking, such as impaired controlled processing (Jennings & Jacoby, 1993; Hay & Jacoby, 1999), slowing in processing speed (Salthouse, 1996), or fewer available processing resources (Craik, 1983, 1986). Altogether, it seems that there are many obstacles preventing optimal recall that become more apparent as one ages. Most of these mentioned deficits hold some sort of implication for the ability to form and recall chunks, though there are few predictions regarding the number of chunks that can be maintained or the size of those chunks. What might this imply for chunk capacity limits and chunk size in older adults, specifically in working memory?

Aging and Chunking

Some prior studies suggest that there may be no age differences in the ability to form chunks out of meaningless strings of items, even if influential grouping is involved (Allen & Coyne, 1988, 1989; Allen & Crozier, 1992). Many of these studies, however, contained a small number of stimuli (e.g., six letters or numbers) that may not greatly tax working memory or provide a need to chunk items together. When stimuli can be grouped more meaningfully, older adults can chunk with greater ease and recall more information, but their improvement in recall is still not as large as the improvement for younger adults (Taub, 1974).

Naveh-Benjamin et al. (2007) examined rates of chunk access and chunk completion in pre-grouped word pairs in young adults, young adults with divided attention (DA), and older adults. A chunk was accessed when at least one item of a word pair was recalled, and chunk completion was the probability that the remaining word in a pair would be recalled when one word of the pair was correctly recalled. Ideally, these measures could determine the number and size of chunks recalled. By these measures and a derived model, young adults recalled considerably more chunks than older adults and recalled somewhat larger chunks (Naveh-Benjamin et al., 2007).

Although Naveh-Benjamin et al. (2007) provide strong evidence for a changing working memory capacity as an effect of aging, there are important issues that remain to be resolved. First is the issue of using unrelated word pairs as stimuli. As words in a pair had no pre-existing associations with each other, strategy is needed for meaningful encoding and recall. A participant might need to create an association between items where one does not exist, which is what the authors intended. While the word pairs are certainly more meaningful and easy to encode than random shapes or numbers, the stimuli may still be too difficult for older adults with depleted resources for such encoding. Second, the formulas used to obtain estimates of size and number were based on tentative assumptions of when chunk formation could occur. Because of this, a more fine-grained analysis of the number and size of chunks recalled may be warranted.

Aging and Language Processing

As previously mentioned, little is known about developmental differences in chunking of sentence stimuli. More is known about developmental differences in general memory for language. Aging effects are found for comprehension and recall of verbal discourse or text. Craik and Masani (1967) noted that while both older adults and young adults show improvement in recall as discourse increasingly approximates English text, older adults show a relative age decrement (this is also discussed in Kausler, 1994). This may be due to the demands of comprehending and integrating language stimuli into working memory, in addition to chunking or transforming information to facilitate encoding (Craik, Morris, & Gick, 1990). Older adults may have difficulty meeting such demands. While older adults perform as well as young adults on comprehension measures, they show deficits in drawing inferences and detecting flaws in passages (Cohen, 1979) and also have worse verbatim memory than young adults, leading to a greater reliance upon gist interpretations, holistic properties of verbal materials, and surface features of speech or prose (e.g., inflection, pauses) (Backman & Nilsson, 1985; Stine-Morrow, Gagne, Morrow, & DeWall, 2004; Stine & Wingfield, 1987). These age differences imply strategic differences in encoding of language stimuli. Sachs (1967) proposed that shallower properties of language (i.e., grammar, syntax) are temporarily stored until the underlying semantic properties can be obtained. Older adults' greater reliance on holistic linguistic properties when verbatim recall is required suggests impairment in online processing and storage of syntactic and grammatical properties. Young adults show competency in verbatim recall; this may mean that they are able to maintain and store surface linguistic properties for longer than would be required for semantic understanding, thus easing recall performance.

Despite older adults employing greater reliance upon gist interpretations or holistic properties of language, the quality and retention of these interpretations is often poorer than expected (Cohen, 1979; Wingfield, Wayland, & Stine, 1992). Indeed, while they are very capable of separating relevant idea units needed for gist interpretation from irrelevant idea units, as propositions increase in discourse, older adults are more likely to remember information that is not central to a passage (Wingfield & Stine-Morrow, 2000). This may be due to less efficient inhibition of irrelevant information, or alternatively, ineffective online processing and updating of information with reduced working memory capacity, as was implicated in some of the studies above.

Either of the above possibilities may disrupt the process of chunk formation of verbal stimuli, particularly when chunk organization is not predetermined. Findings from language studies discussed in Kemper (2006) appear to support the idea that aging impairs chunking for language stimuli via reduced capacity. When strategies to divide verbal stimuli into smaller segments (or rather, chunks) are examined across younger and older adults, older adults often process smaller segments than young adults. Additionally, older adults segment verbal stimuli in a more haphazard fashion that young adults, often disregarding those boundaries present in stimuli that can serve as markers for dividing language materials into smaller pieces (e.g., sentence boundaries). Supporting this statement, Fallon, Kuchinsky, and Wingfield (2006) presented both younger and older adults with spoken passages containing alternating one-clause and two-clause sentences that abruptly stopped after two, four, or six sentences were spoken, and this interruption always occurred after a two-clause sentence. Participants were instructed to recall the words leading up to the interruption (see Jarvella 1970, 1971 for more details about this paradigm). Using strict verbatim recall, Fallon et al. (2006) found that nearly all participants recalled the clause immediately prior to interruption, suggesting reliance upon a clause structure for comprehension, and recall of one clause was the most common occurrence for both age groups. Aging, however, had an impact upon recall. Older adults were less likely than younger adults to recall two consecutive clauses, and they were more likely than younger adults to recall less than a full clause. In the context of chunking, these may be thought of as a reduced number of chunks, reduced chunk size, or both possibilities. The deficits may be explained by an associative deficit (i.e., linking words together) and by greater difficulty for older adults in meeting the demands of online language processing (here, comprehending the underlying semantic structure as well as retaining the verbatim surface structure) (Fallon et al., 2006). Both Kemper (2006) and Fallon et al. (2006) observed analogs of chunk size estimates across age groups that were similar to chunk size patterns found in Naveh-Benjamin et al. (2007). These findings imply that chunk formation may be disrupted because of poor encoding and grouping strategies in addition to an impaired working memory capacity.

Current Study

The current study involved two important components that should provide an extension of the findings of Naveh-Benjamin et al.(2007). First, English sentences were used as stimuli in this study, which may ease task difficulty due to the syntactic and semantic associations that are often present in natural language, as older adults are often helped by pre-existing or predetermined associations (Hay & Jacoby, 1999; Naveh-Benjamin, Craik, Guez, & Kreuger, 2005).

Second, a response-timing analysis was used for a more microscopic investigation of chunk boundaries. Similar to research on the reconstruction of chess games from immediate memory (e.g., Chase & Simon, 1973; Gobet & Clarkson, 2004) and the short-term serial recall of lists (Anderson & Matessa, 1997; Maybery, Parmentier, & Jones, 2002), the assumption was that the slow process of retrieval of a chunk from long-term memory is followed by the faster process of recalling the items within a retrieved chunk, leading to pauses between chunks. The scoring of recall estimates chunk boundaries by making use of pauses in conjunction with the assumption of coherence in language (Tulving & Patkau, 1962) and compares these estimated chunks with the measures adopted by Naveh-Benjamin et al. (2007).

Clearly chunking is not the only influence on recall; another influence is the length of stimuli. Use of sentences may offset possible impairments in recall due to an increased amount of information (i.e., words, morphemes, and phonemes) per chunk, and presenting items in comprehensible language may allow for a better understanding of how chunks and associations are grouped and formed in recall, as well as how these might change with age. Additionally, though, by presenting sentences of varying lengths, it was possible to assess the role of time, which should be important according to the phonological loop account conceived by Baddeley (1986) for material within the limit of the loop. Baddeley, Thompson, and Buchanan (1975) found that recall for verbally presented items was generally limited to what participants could rehearse in 2 s, providing support for a time-limited capacity. If capacity limits in working memory are due to time constraints within a phonological loop, it is expected that recall for sentences exceeding this time limit should be very poor; otherwise, an alternative account for these capacity limits will be necessary.

As previously mentioned, older adults generally recall fewer chunks from working memory than young adults, and the chunks recalled are often smaller in size. One aim of this study involved further investigation of the effects of aging upon qualitative and quantitative aspects of chunking (e.g., number of chunks, chunk size). The critical manipulation for this study involved the grouping of the stimuli. Participants were presented with one of four types of lists: (1) lists of four random "sentences," each comprising words from various sentences mixed together and randomly ordered, as a control condition; (2) lists of 4 short sentences (e.g., *Then she walked to school; My aunt cooked chicken; I smashed the vase; The woman heard footsteps*); (3) lists of eight such short sentences; and (4) lists of four long sentences made of two meaningfully conjoined short sentences (e.g., *turn the block over and place it by the toy*). As shown in Table 1 and Appendix A, no participant received the same content words in more than one condition.

The purpose of long sentences was to determine whether sentence recall operates under limits of chunk size or length. It was compared to the two types of trials containing short sentences. The 4-short-sentence trials were, in principle, equivalent to 4-long-sentence trials in terms of the number of chunks presented, with 4 sentences each. In contrast, the 8-short-sentence trials had twice as many sentences as the 4-long-sentence trials, but were was equivalent to them in terms of number of words presented or overall length of phonological material within the trial. Recall for each of these trials can be examined to determine whether the limit in performance is in the number of chunks that can be recalled or in the length of speech that can be recalled. Of course, a chunk limit can predict an intermediate result if the pairs of phrases within each of the four long sentences are not perfectly conjoined to form longer sentences in the memory representation. Similarly, the lists of random words and the lists of four short sentences both are equivalent in terms of the length of speech, but the lists of random words contain many more chunks, inasmuch as words rather than sentences are the coherent units likely to comprise chunks.

To best understand what and how much was being stored in working memory, the task was designed to prevent rehearsal of sentences after they were presented by using a relatively short (1-s) interstimulus interval between the spoken sentences. In any case, the lengths of sentences presumably preclude the use of Baddeley's 2-s loop for more than 1 sentence.

It is possible to anticipate patterns of responding based on different recall principles. If each sentence becomes a chunk and the mean capacity in young adults does not exceed 4.0 then, according to a capacity-limit hypothesis, there should be an aging deficit in all conditions. In every condition, the older adults should recall fewer chunks than young adults.

A different prediction based on a chunk capacity limit is possible if participants are able to hold in mind a certain limited number of associations between sentence units, forming still larger chunks online. For example, they might make up a novel scenario that logically links two unrelated sentences together. If so, an age difference in the number of chunks recalled might only emerge when the number of chunks to be recalled considerably exceeds four chunks. In such conditions there is a much larger need to associate chunks in order to facilitate recall. This is the case at least in the 4 random sentences condition and in the 8 short sentences condition. For an example of what the data might look like based on this capacity-related process, see Figure 1.

The predictions of an associative deficit hypothesis support the above suggestion. Older adults have pronounced difficulty in binding multiple objects or features together in recall (Mitchell et al., 2000; Cowan et al., 2005). It is possible that in this task, multiple types of binding may have been involved. There may be bindings within constrained sentences (i.e., short sentences, one sentence within a long sentence), bindings between two short sentences comprising one long sentence, and those between unassociated words (as in four random sentence trials) or sentences. Beyond binding stimuli, there may also be binding between a given stimulus and temporal context. While, as mentioned below, there were no instructions for serial recall, there must be a temporal marker for each sentence that is presented so that one may know which items must be recalled, as well as at what time. Whereas lists with only four chunks can be held mostly as separate chunks (though it should depend on the participant's capacity), lists with a greater number of chunks present a greater need to form new associations on line to allow retention of a larger proportion of the items. Therefore, as shown in Figure 1, older adults' inability to do so would cause much larger deficits in the four random and eight short sentence conditions than in the other conditions.

In contrast to the above possibility, if only a time-limited working memory capacity influenced chunk formation and recall, then the overall number of words present (and hence, the overall duration) in a given trial condition would be a primary factor in performance. In this case, older adults might perform comparably to young adults only on those trials where the overall number of words presented and the total duration of that trial were relatively small (i.e., four random and four short sentences). If the total number of words and duration in a trial condition were larger, as it is for four long and eight short sentences, older adults should show impairment relative to young adults on those trials (see Figure 2).

Each of the above factors can produce unique effects and interactions between age and trial condition. It is also possible that more than one of the above factors may have influenced the results of this study. For example, a combination of the factors discussed with respect to Figure 1 and Figure 2 together provide reasons to expect an aging deficit in every condition with the possible exception of four short sentences, as shown in Figure 3 Other combinations of factors are of course possible, but they are not discussed here.

Method

Participants

Participants were 42 University of Missouri undergraduates (25 males, 17 females) and 25 older adults (7 males, 18 females) recruited from the Columbia, Missouri community, and all were randomly assigned to testing groups. Data from two young adults and one older adult were excluded due to experimental error. To ensure an equal gender ratio across age groups, data from the first seven males and 17 females were used in the young adult group, creating two age groups with 24 participants each. Young adults had a mean age of 18.37 years and a mean education level of 13 years; the mean age for older adult participants was 70.75 years, with a mean education level of 14.04 years. Undergraduates received course credit in exchange for participation, while older adults received monetary compensation.

Materials

Stimuli were created through the computer program Audacity. These sentences were verbally presented in a female voice through headphones from 45 to 70 dB. Both the experimental procedure itself and recording of sound data from participants were administered through the computer software program Revolution 2.6.1, which also provided a graphical interface for the experiment in addition to randomizing and presenting sentence stimuli. Participants responded by speaking through a microphone. Responses were converted into sound files and saved by the computer program.

Design

Spoken sentences included words taken from age of acquisition norms via the MRC psycholinguistic database (http://www.psy.uwa.edu.au/mrcdatabase/uwa_mrc.htm), with acquisition ratings from 100 to 350 serving as the parameters. Trials contained short, long, or random sentences. Short sentences contained two words from the norms, and were 3 to 5 words long (e.g., “Don’t scare your brother”). Long sentences were composed of two short sentences that were connected, and were 8 to 11 words long (e.g. “Don’t scare your brother because the screaming bothers me”). Random sentences were similar to simple sentences in that each contained two words from the norms and were 3 to 5 words long. However, the words were in a non-grammatical, haphazard order (e.g. "aunt jersey cooked almost").

Four different conditions were formed to compare the number of chunks and the total length of those chunks (here, the number of words) across stimuli types. The material to be remembered in a trial included one of the following: (1) 4 short sentences, (2) 4 long sentences, (3) 8 short sentences, or (4) 4 random sentences. For example, 4 long sentences and 8 short sentences will be equivalent in overall length; in terms of the number of chunks, 4 long sentences will be equivalent to 4 short sentences if perfect chunking is assumed. There were two trials in each of these four conditions, and the order of these eight trials was randomized across participants.

Two groups were created, with each group receiving counterbalanced sentences across condition. The sentences used for short and long sentence trials were equated across subjects, so any differences in performance that occurred could not be attributed to the selection of materials. Sentences that were a part of long sentence conditions in the first group were presented as short sentences in the second group. As an example, participants in Group 1 were presented with the long sentence "Play in the park but expect rainy weather later." For participants in the second group, this sentence was deconstructed into two smaller sentences ("Play in the park" and "Expect rainy weather later"), and these were placed into either the four or eight short sentences condition. In this case, these sentences were placed into the eight short sentence conditions. In the same vein, long sentences that were presented to the second group were deconstructed into short sentences to be presented to the first group. Short sentences that would otherwise be connected in a long sentence condition were placed in different short sentence conditions to reduce sentence-specific effects. Using the above example, “Play in the park” was placed in the first trial of the eight short sentences condition; “Expect rainy weather later” was placed in the second trial of eight short sentences. In a similar fashion to the above method for counterbalancing with long sentences, sentences that were presented as short sentences could be combined to form a long sentence, or combined with other short sentences to form random sentences. For example, the sentence “We took pictures as a gift” was deconstructed and combined with other short sentences to form the random sentences “the struck pictures your sting” and “gift building a as” for Group 2. An illustration of how sentences were reassigned to different trial conditions to achieve counterbalancing is included in Table 1; for additional details on the stimuli, trials, and presentation order, refer to Appendix.

Procedure

All participants were tested one at a time, with young adults in a sound-attenuated booth and older adults in a quiet room; external noise was reduced for both testing sites. Both locations were equipped with a computer to run the experiment program, headphones for listening to stimuli, and a microphone for recording spoken responses. After completing necessary informed consent and demographic forms, participants were instructed to listen carefully to the stimuli and, when cued, verbally recall what they had just heard in any order. Each of the eight trials began with a 1000-ms fixation on the computer screen (the word “Ready”) to alert participants. The spoken stimuli were then presented through the headphones. Each sentence in a trial condition was separated by a pause of 1000 ms. At the end of the set of sentences and a final pause of 1000 ms, a 500-ms, 400 Hz tone was presented as the cue for participants to verbally recall the stimuli. Participants provided their answers by speaking through a microphone. This free recall period had a maximum duration of 1 minute, though participants were free to terminate recall before that time by pressing a key. Another key press started the next trial. All responses were converted into sound files and saved by the computer program.

Analyses

All participant responses were verbally transcribed by the experimenter. For additional analysis, responses were carefully timed using the Audacity program which provides an oscillographic display along with selective auditory playback to allow accurate millisecond timing. As participants could recall words or sentences in any order, duration of sentences or comprising words that made up sentences from each condition was recorded. The timing information contributed to an adopted-chunks measure to be described below.

Six different measures were used to examine age differences in recall of the different trial conditions: total words recalled, adopted chunks, chunk size, sentence access, sentence completion, and serial position. *Total words recalled* simply involved the number of words correctly recalled from each condition. If more than one instance of a word was present within a trial (e.g., “the” appeared twice), each occurrence in recall was counted as a separate word recalled. In this example, if “the” was recalled twice, two words were recalled. *Adopted chunks*, a measure developed by Tulving and Patkau (1962), was used here in a modified form that took pauses into account. An adopted chunk was defined as any series of words that was recalled with the words in the presented order and without any internal pause of 2 s or longer (a duration used to identify chunks in studies of spatial memory, e.g., Chase & Simon, 1973; Gobet & Clarkson, 2004). Based on these criteria, an adopted chunk could be as short as one word, functioning as the smallest possible unit that can be increased through additional binding or recoding, or as long as the entire trial condition itself. *Chunk size* was calculated on the basis of the previous two measures, as the number of words recalled per adopted chunk. *Clause access* referred to access to the coherent sentential information provided, and was measured as the number of clauses in a trial from which at least one word was recalled. In the case of trials containing four long sentences, each long sentence was made up of two smaller clauses, resulting in eight total clauses that could be accessed for that trial condition; for four random sentences, each random sentence (as opposed to each word) was counted as a clause, with a maximum of four clauses that could be accessed. Similar to the measure of chunk completion used by Naveh-Benjamin et al. (2007), *clause completion* was measured as the proportion of words recalled for a given clause on the condition that the clause had previously been accessed. These measures were followed by an analysis of *serial position* to determine whether young and older adults displayed different patterns of access and word recall across the different serial positions in a given trial. This may provide insight into how the different age groups strategically organize the sentences for later recall.

Results

The means and standard errors for the measures explained above were calculated for each age group across trial condition, and can be found in Table 2. The results for each of the measures will be discussed in turn.

Total Words Recalled

A repeated-measures ANOVA for total number of words recalled included age group as a between-groups factor and trial type as a within-group factor. There was a significant main effect of age group, $F(1, 46) = 9.03$, $\eta_p^2 = .164$, $p < .05$, with young adults recalling more words (*M* = 10.81, *SD* = 2.48) than older adults (*M* = 8.66, *SD* = 2.48). There was also a significant effect of trial condition, $F(3, 138) = 112.12$, $\eta_p^2 = .709$, $p < .05$, with the greatest number of words recalled for four long sentences (*M*= 14.02, *SD* = 4.52), followed by four short (*M* = 10.52, *SD* = 2.68), eight short (*M* = 9.67, *SD* = 3.52), and four random sentences (*M* = 4.73, *SD* = 1.83). These effects were qualified by a significant age group x trial interaction, $F(3,138) = 4.23$, $\eta_p^2 = .084$, $p < .05$. As Figure 4 suggests, as the number of words increased in sentences, and as the amount of coherence increased, recall performance was generally higher (i.e., number of words recalled was highest for four long sentences and lowest for the condition termed four random sentences, which are half as long and have no linguistic coherence). Additionally, young adults appear to have generally recalled more words than older adults.

Post-hoc Newman-Keuls tests provided more information about the above findings. For young adults, recall followed the pattern: four long > four short = eight short > four random. For older adults, the pattern was slightly different: four long > four short > eight short > four random. That is, four short and eight short sentences produced equivalent numbers of words recalled in young adults, but the larger number of sentences diminished recall in older adults. Young adults recalled roughly the same number of words from both conditions, which may indicate a capacity limit. In contrast, older adults recalled more information from four short than from eight short sentences, suggesting that adults may not be able to avoid interference from additional material after their capacity limit has been exceeded.

The only significant post-hoc comparison between age groups for a given condition was found for four long sentences ($p < .05$). This finding appears to confirm some of what can be seen in Figure 4, namely that there was no difference in recall for four short sentences and a large difference in recall across groups for four long sentences. The magnitude of the difference in mean words recalled for eight short and four random sentences may not have been large enough to approach significance with these tests. Independent-sample t-tests did confirm significant differences in words recalled across age groups (four long, $t(46) = 2.99$, $p < .05$; eight short, $t(46) = 2.47$, $p < .05$, and four random, $t(46) = 3.5$, $p < .05$).

Adopted Chunks

A repeated-measures ANOVA for mean adopted chunks (the measure modeled after Tulving & Patkau, 1962, but with inter-word pauses also taken into account) included age group and trial condition as between and within-participant factors, respectively. There was a significant main effect of age group, $F(1,46) = 7.26$, $\eta_p^2 = .136$, $p < .05$, with young adults forming more chunks (*M* = 2.76, *SD* = .609) than older adults (*M* = 2.28, *SD* = .609). There was also a significant effect of trial condition, $F(3,138) = 13.6$, $\eta_p^2 = .228$, $p < .05$, with the largest number of chunks formed for eight short sentences (*M* = 2.88, *SD* = 1.17), followed by four short (*M* = 2.71, *SD* = .81), four long (*M* = 2.56, *SD* = .78), and four random sentences (*M* = 1.92, *SD* = .82). There was no significant age x trial interaction, $F < 1$. As seen in Figure 5 older adults show a deficit in the number of chunks formed relative to young adults, replicating the findings of

Naveh-Benjamin et al. (2007). Additionally, across age groups, chunks were more difficult to form when participants were presented with stimuli that were not linguistically coherent (here, four random sentences). Figure 5 suggests that as long as the language stimuli are coherent, there should not be differences in the number of chunks that are formed, which may indicate a capacity limit.

Newman-Keuls tests confirmed the above findings, though imperfectly. Young adults displayed the following patterns: four short = eight short = four long, four short and eight short > four random, and four long = four random. While adopted chunks were easier to form for those stimuli that were coherent, young adults had greater difficulty forming chunks for longer sentences. Older adults showed a different pattern, four short = eight short = four long > four random. Confirming the lack of interaction, no significant age comparisons were found for any sentence condition. These findings suggest that, particularly for older adults, more chunks are formed for coherent sentences than sentences that are not coherent; chunks formed from coherent stimuli are not dependent upon the length of the sentence, as the number of chunks formed does not increase as the number of words in a sentence increases.

Chunk Size

In terms of chunk size, analyses using a repeated-measures ANOVA confirmed the lack of a main effect of age group, $F < 1$, in contrast to the significant main effects of age for both words recalled and chunks recalled. There was a significant effect of trial condition, $F(3,135) = 53.0$, $\eta_p^2 = .54$ $p < .05$, with the largest chunks produced with four long sentences (*M* = 5.68, *SD* = 1.78), followed by four short (*M* = 4.13, *SD* = 1.38), eight short (*M* = 3.49, *SD* = .89), and four random sentences (*M* = 2.64, *SD* = 1.05). The interaction between age group and trial was not significant. As Figure 6 shows, there does not appear to be any aging effect upon chunk size, as the means across age groups for each trial condition appear comparable. This is contrary to the moderate aging deficit found by Naveh-Benjamin et al. (2007). Across both groups, the mean chunk size varied across sentence condition, with its highest value for four long sentences and its lowest value for four random sentences. These findings appear reasonable, given that the measure of chunk size depends on the number of words recalled and the number of chunks formed. As found above, recall of words and chunk formation are facilitated when sentences contain additional information and are coherent. Newman-Keuls tests generally confirmed this, with young adults showing the pattern four long > four short = eight short > four random. Older adults displayed a pattern of four long > four short > eight short > four random. As was seen for recall of total words, older adults displayed a larger estimate for four short sentences than for eight short sentences; young adults had comparable chunk size estimates for both conditions. This finding is likely due in part to the measurement of chunk size, which takes the total number of words into account.

Clause Access

For the following two measures, like Naveh-Benjamin et al. (2007), we define each chunk as the units presented (here, clauses, or isolated words in the random condition), rather than determining the chunk boundaries empirically. For clause access, a repeated measures ANOVA produced a significant effect of age group, $F(1, 46) = 12.40$, $\eta_p^2 = .212$, $p < .05$. Young adults accessed a greater number of clauses (*M* = 3.27, *SD* = .65) than older adults (*M* = 2.60, *SD* = .65). There was also an effect of trial condition, $F(3, 138) = 81.40$, $\eta_p^2 = .638$, $p < .05$, with the greatest number of clauses accessed for four long sentences (*M* = 4.08, *SD* = 1.12), followed by four short (*M* = 3.02, *SD* = .63), eight short (*M* = 2.84, *SD* = 1.05), and four random sentences (*M* = 1.81, *SD* = .69).These effects were qualified by a significant age group x trial interaction, $F(3, 138) = 3.70$, $\eta_p^2 = .074$, $p < .05$. Consistent with the measure of the number of chunks recalled, Figure 7 suggests that young adults generally accessed more clauses than older adults. The access deficit in older adults appeared particularly pronounced for those cases where the number of sentences presented or the length of presented sentences is long (i.e., eight short sentences, four long sentences, and four random sentences due to their lack of coherence).

Newman-Keuls tests suggested a difference between age groups in terms of access. First, a significant difference between age groups was found for four long sentences, as was found for mean words recalled. Second, young adults displayed the pattern of four long > four short = eight short > four random. Older adults showed a different pattern: four long > four short > eight short > four random. Both younger and older adults are able to take advantage of the overarching semantic and linguistic structure in long sentences, leading to higher access rates than for the shorter sentences. Regarding short sentences, while young adults showed similar access rates for the four short and eight short sentence conditions, older adults displayed differential rates of access. This difference between age groups suggests that older adults may be more sensitive to the total number of words presented in a trial, while young adults may be more sensitive to the chunks denoted by sentence composition.

Clause Completion

As previously mentioned, clause completion was measured as the proportion of words recalled from a given clause, provided that it had already been accessed. A repeated measures ANOVA found no significant effect of age upon mean completion, $F<1$, contrary to the findings of Naveh-Benjamin et al. (2007). Consistent with this finding, Figure 8 shows no apparent differences between age groups on completion rates for any sentence condition. There was a main effect of trial type, $F(3,111) = 17.22$, $\eta_p^2 = .317$, $p <.05$. The greatest proportion of clause completion occurred for those trials containing eight short sentences (*M* = .829, *SD* = .163), followed by four short sentences (*M* = .813, *SD* = .107), four long sentences (*M* = .727, *SD* = .124), and four random sentences (*M* = .649, *SD* = .199) respectively. The interaction between age group and trial condition was not significant, $F<1$. Post-hoc Newman-Keuls tests revealed that completion rates were not significantly different between four short and eight short sentence trial conditions, resulting in the overall pattern eight short = four short > four long > four random. When separated on the basis of age group Newman-Keuls tests revealed slightly different patterns of mean clause completion. Older adults showed the same pattern found above. In contrast, young adults displayed the pattern eight short = four short = four long > four random. It appears that older adults' completion rates may be more sensitive to the length of coherent sentences, as average completion for long sentences was lower than completion of short sentences, while this discrepancy was not present for young adults.

Serial Position Effects

Given the results above, we were interested in whether the measures that yielded aging effects were due to deficits in strategically linking clauses together, whether part of a superordinate sentence or not. One possible way of examining this was by looking at serial position effects for those measures where aging was a significant factor. Because adopted chunks and chunk size could span multiple clause boundaries despite producing significant effects of aging, we examined serial position effects for clause access and words recalled.

Repeated measures ANOVAs were run separately for each trial type for clauses accessed. For four long sentences, there were main effects of age, $F(1, 46) = 11.94$, $\eta_p^2 = .206$, $p <.05$, as well as for serial position, $F(7, 322) = 21.99$, $\eta_p^2 = .323$, $p <.05$, but there was no significant interaction. Young adults accessed clauses at a higher rate (*M* = .58, *SD* = .14) than the older adults (*M* = .44, *SD* = .14), and clauses at the end of a trial were more likely to be accessed (eighth position, *M* = .86, *SD* = .27; seventh, *M* = .76, *SD* = .31) than clauses in earlier positions (first, *M* = .57, *SD* = .37; second, *M* = .27, *SD* = .29; third, *M* = .40, *SD* = .33; fourth, *M* = .36, *SD* = .30; fifth, *M* = .51, *SD* = .37, and sixth, *M* = .33, *SD* = .31). Post hoc Newman-Keuls tests showed that access for those clauses part of the last sentence (i.e., seventh and eighth serial positions) have significantly different rates of access from all other clauses; as Figure 9 suggests, their access rates are much higher. Additionally, the tests showed that older adults were particularly sensitive to clauses that began long sentences, but were less likely to access the following clause that comprised that particular sentence. Access for the first clause was significantly higher than the second clause (making up the first sentence). This was not found for the two sentences that followed, despite the appearance of a similar pattern for those sentences in pre-recency positions.

Similar results were found for both four random and eight short sentences. Both trial types yielded main effects of age (four random, $F(1, 46) = 9.74$, , $\eta_p^2 = .174$, $p <.05$; eight short, $F(1, 46) = 6.37$, $\eta_p^2 = .121$, $p <.05$), and young adults accessed more clauses (four random, *M* = .53, *SD* = .17; eight short, *M* = .40, *SD* = .13) than older adults (four random, *M* = .37, *SD* = .17; eight short, *M* = .30, *SD* = .13). There were also significant serial position effects (four random, $F(3, 138) = 31.21$, $\eta_p^2 = .404$, $p < .05$; eight short, $F(7, 322) = 25.58$, $\eta_p^2 = .357$, $p < .05$) for both trial types. For four random sentences, access rates were higher for the sentence in the last serial position (*M* = .81, *SD* = .28) than for sentences in the remaining serial positions (first, *M* = .22, *SD* = .19; second, *M* = .32, *SD* = .26, and third, *M* = .44, *SD* = .24). A similar recency effect was found for eight short sentences, with sentences presented near the end of a trial having higher access rates (eighth position, *M* = .85, *SD* = .25) than sentences in earlier serial positions (first, *M* = .34, *SD* = .33; second, *M* = .23, *SD* = .29; third, *M* = .14, *SD* = .26; fourth, *M* = .27, *SD* = .34; fifth, *M* = .27, *SD* = .30; sixth, *M* = .23, *SD* = .28, and seventh, *M* = .47, *SD* = .40). Both analyses, along with the analysis for four long sentences, failed to find a significant interaction between age group and serial position (four random, $F < 1$; eight short, $F(7, 322) = 1.52$, $p > .05$); this suggests that while young adults accessed a larger number of clauses, there were no significant differences between age groups in patterns or strategies of access across serial positions. Access rates for four long sentences were generally higher across serial positions than for eight short sentences, even though both trials contained the same number of clauses. This suggests that, similar to Taub's (1974) findings, increased coherence and meaningfulness of stimuli lead to improvements in access.

Analyses for serial position for four short sentences yielded unexpected findings. There was neither a significant effect of age, $F < 1$, nor of serial position, $F(3, 138) = 2.13$, $p > .05$. As with the other trial conditions, there was no significant interaction between age and serial position, $F(3, 138) = 1.08$, $p > .05$. In the case of four short sentences, there appeared to be no evidence that either aging or serial position affected access of clauses; additionally, there did not appear to be any strategic difference in access of clauses in their respective serial positions across age groups.

Similar repeated measure ANOVAs were used to examine the frequency of words recalled across serial positions in a trial; like the analyses for access, each of the four trial types was analyzed separately. Analyses produced a pattern of findings that were similar to the findings for clause access and serial position. Main effects of age were found for four long, four random, and eight short sentence conditions (four long, $F(1, 46) = 6.51$, $\eta_p^2 = .124$, $p < .05$; four random, $F(1, 46) = 10.37$, $\eta_p^2 = .184$, $p < .05$, and eight short, $F(1, 46) = 7.33$, $\eta_p^2 = .137$, $p < .05$). In all cases, young adults recalled a larger number of words across all serial positions (four long, *M* = 1.70, *SD* = .52; four random, *M* = 1.34, *SD* = .45, and eight short, *M* = 1.33, *SD* = .42) than older adults (four long, *M* = 1.32, *SD* = .52; four random, *M* = .92, *SD* = .45, and eight short, *M* = 1.00, *SD* = .42) The number of words recalled per clause was also significantly affected by serial position for these trials (four long, $F(7, 322) = 31.38$, $\eta_p^2 = .405$, $p < .05$; four random, $F(3, 138) = 55.30$, $\eta_p^2 = .571$, $p < .05$, and eight short, $F(7, 322) = 31.91$, $\eta_p^2 = .409$, $p < .05$). In all cases, more words were recalled for those clauses in the final serial position of a trial (four long: eighth position, *M* = 2.94, *SD* = 1.29; four random: fourth position, *M* = 2.72, *SD* = 1.31, and eight short: eighth position, *M* = 3.12, *SD* = 1.11) than for clauses in earlier serial positions (four long: first, *M* = 1.50, *SD* = 1.14; second, *M* = .81, *SD* = .95; third, *M* = 1.18, *SD* = 1.08; fourth, *M* = .93, *SD* = .97; fifth, *M* = 1.28, *SD* = 1.22; sixth, *M* = .75, *SD* = .83; seventh, *M* = 2.69, *SD* = 1.30; four random: first, *M* = .40, *SD* = .63; second, *M*

= .72, *SD* = .98; third, *M* = .66, *SD* = .64; eight short: first, *M* = 1.09, *SD* = 1.11; second, *M* = .62, *SD* = .85; third, *M* = .45, *SD* = .86; fourth, *M* = .83, *SD* = 1.06; fifth, *M* = .86, *SD* = 1.05; sixth, *M* = .61, *SD* = .82; seventh, *M* = 1.75, *SD* = 1.66). Like with the access measures, none of these trials showed a significant interaction between age group and serial position (four long, $F(7, 322) = 1.46$, $p > .05$; four random, $F < 1$; eight short, $F(7, 322) = 1.42$, $p > .05$), which suggested that there was no strategic difference across age groups in recalling words across different serial positions (see Figure 10). Similar to the pattern found for clause access for four long sentences, older adults recalled a greater number of words from a clause that began a long sentence than the following clause that completed the sentence, as long as the clauses were not in recency positions. However, post hoc tests showed that only the first and second serial positions (clauses making up the first sentence) had substantial differences in words recalled to be statistically significant.

As was found for access across serial positions, analyses failed to find significant effects of age, $F < 1$, or serial position, $F(3, 138) = 1.49$, $p > .05$, in words recalled per clause for four short sentences. In addition to failing to find significant main effects, we also failed to find a significant interaction between age group and serial position, $F < 1$. The lack of differences between age groups for this particular trial for both the frequency of words recalled and rate of clause access is difficult to understand in light of our hypotheses, and will be discussed in detail below.

Discussion

Several findings from this study provide insight into which types of language stimuli young and older adults will be more likely to recall, as well as where aging deficits might be found using the various measures described above. It was expected that sentences that were more meaningful in terms of their coherence would be better recalled than those sentences that had little or no coherence. We found that young and older adults were indeed more likely to recall words from these meaningfully constructed sentences than pseudo-sentences that were haphazard in their structure. Furthermore, when participants were presented with the former, they tended to access a greater number of clauses, form a larger number of chunks, had larger chunk size estimates, and were more likely to complete a previously accessed clause. This is consistent with previous studies where increases in stimuli meaningfulness lead to improvement in chunk formation and recall (e.g., Tulving & Patkau, 1962; Glanzer & Razel, 1974). Interestingly, connecting short sentences together to increase total number of words and linguistic connections within one large sentence improved performance on certain measures above the level of short sentences, namely words recalled, chunk size, and clause access. Connecting two short sentences together allowed participants to increase the amount of information present within one chunk. While combining two short sentences into one larger sentence provided some benefits above short sentences, these improvements in performance were never doubled in size; this suggests that long sentences may increase chunk size, but at a diminishing rate, possibly due to reaching or even exceeding working memory capacity.

Additionally, the data provides evidence against a purely time-limited account of working memory capacity limits, proposed by Baddeley (1986). For many measures collapsed across age groups, best illustrated by the total number of words recalled in a trial, performance on trials longest in duration (four long and eight short) were often either comparable or better than the level of performance for four short sentences, where the duration was much shorter. Furthermore, performance for four random sentences is always worse than four short sentences in all measures examined, despite having similar durations. If working memory is only limited by time constraints within a phonological loop, the information that can be stored and recalled from trials with longer duration cannot be greater than that for trials shorter in duration; conversely, trials that have similar duration should have comparable performance levels. While a length-based working memory capacity limit may influence storage, grouping, and recall, it cannot be the sole influence in this study. A chunk-limited capacity, in contrast, is more reasonable given the results. Here, perfect chunking does not occur—as previously mentioned, performance levels for long sentences are not two times larger than for short sentences. However, this account of capacity limits is able to account for differences in performance across trial types based on a sentence's ease of grouping and its composition (i.e., number of words, semantic and syntactic connections). It is also able to account for instances where performance levels for eight short sentences are comparable to those of four short sentences; in these cases, the short sentences have a similar linguistic composition and therefore can be stored in working memory and recalled in a similar manner. While the chunk capacity account is better able to explain the results across all trials, it is unclear whether this is the only influence on recall in this study.

Although increased meaningfulness of sentences generally improved performance on all measures examined, the results also showed that there are clearly aging deficits in chunking and recall, consistent with studies by Craik and Masani (1967) and Taub (1974). While some findings replicated previous research on aging and chunk formation, other results unexpectedly did not. As one certainly would expect, relative to young adults, older adults recalled fewer words, both overall and across various serial positions within a trial. Also, older adults formed fewer adopted chunks, as had previously been found. However, in contrast to what was observed by Naveh-Benjamin et al. (2007), the size of the chunks that older adults formed were, overall, no smaller in size than those of young adults. Similarly, although older adults had lower rates of clause access than young adults, the former group was no less likely than the latter to complete clauses that had already been accessed. The discrepancy from Naveh-Benjamin et al. could be explained on the grounds that the items to be bound together were primarily sentence-based rather than being random word lists. Even in the case of four random sentences, the parts of speech are mixed and perhaps it was possible for participants to perceive each random sentence as an oddly-structured message of some sort.

In examining the trial types separately, many of the aging deficits that were found were primarily due to performance on four long, four random, and eight short sentence conditions. In all measures examined, older adults always had lower performance levels relative to young adults on these trials, though the differences between age groups were not always significant in post hoc tests. There are different possibilities as to why these deficits may have occurred. First, older adults may have been more sensitive to the number of items presented as opposed to what comprised the sentences themselves, suggesting an increased reliance upon length limits in working memory capacity. As was previously found, chunk limits and length limits may operate differentially in working memory (Chen & Cowan, 2005), and perhaps aging produces a greater reliance upon length limits, especially for particularly long lists of items. Many of the post hoc tests used on the above measures showed that while young adults showed comparable performance between four short and eight short sentences, older adults always had higher performance for four short sentences than for eight. This may provide some evidence that older adults are more affected by the total length of the trial condition, while young adults may be more sensitive to the sentences and composition as opposed to their length. While this is possible, given the differences between young and older adult performance for four long and eight short sentences, it cannot fully account for the aging deficits that were found. Even for older adults, recall and clause access was always higher for the connected long sentences than for eight short sentences (consistent with Taub, 1974), despite having similar overall length. Furthermore, older adults displayed deficits relative to young adults for four random sentences; however, this trial type was considerably shorter than either the four long or eight short conditions, conditions where performance was much higher.

Second, older adults may have utilized a different means to encode and group sentences into working memory. Findings from the serial position analyses show that this is probably not the case. Given the nature of the results, it is more reasonable to suggest that while older adults access and recall less than young adults from each serial position, their means of organization, with a large reliance upon recency, are not qualitatively different.

Third, older adults may have a reduced working memory capacity, with fewer sentences stored overall. This seems plausible, given the aging deficits seen for four long, four random, and eight short sentences. However, an appeal to capacity limits cannot fully account for their comparable performance to young adults for four short sentence trials on all measures examined. If a reduced capacity were to fully account for all age deficits, older adults would have had lower performance relative to young adults on all trials. While differences in working memory capacity limits are likely a strong influence on what is stored, chunked, and eventually recalled, the age deficits found are not due to this factor alone.

Finally, older adults may have shown deficits in performance because they encountered more difficulty binding words and sentences together, as the associative deficit hypothesis would predict. This may certainly explain the large differences in performance between the two age groups for four long and four random sentences, two trial conditions where more associations would have to be formed. We were particularly interested in whether older adults would be able to bind two short sentences that compose a longer sentence, and expected that associative deficits would make this demand difficult for older adults to meet. Clause completion provides the best evidence either for or against binding words and connections together into a cohesive sentence; for four long sentences, no age differences in completion rates were found. This suggests that, at least for individual clauses, older adults were as competent as young adults in binding words together into a cohesive short sentence. To examine how both age groups bind two clauses together into one long sentence, as well as how two separate long sentences might be bound together, findings on serial position provide insight. For the former, access for two clauses that are part of a long sentence was typically higher than access for two individual short sentences (as seen with eight short sentences). Furthermore, serial position figures indicated that older adults were sensitive to those clauses that began sentences, particularly for those sentences presented first. There did not appear to be an aging deficit for this form of binding, as the differences between access and word recall for the two clauses in a sentence were not significant and young adults displayed a similar pattern. An aging deficit in binding might have been found for the latter aforementioned possibility (i.e., binding separate long sentences). Although young adults displayed substantial dips in access for given serial positions, their access rates for clauses across all serial positions were far more consistent than those of older adults. More consistent access may suggest that young adults were better able to group different long sentences together, despite weak associations between them. Older adults, on the other hand, may have encountered more difficulty in forming such novel associations. Thus, though not in the manner expected, the associative deficit influenced the aging deficits present in our findings. This deficit, like the other influences on chunking discussed above, was not the sole influence upon our findings, as four short sentences showed no significant differences between age groups on any of the measures obtained. As none of the possible factors proposed can account for the findings alone, it is likely that each of these has a degree of influence upon age differences in forming and recalling chunks, though the extent to which each is involved cannot be determined from the measures used. Additionally, other factors that were not discussed here may be involved in this task (e.g., long-term memory components), and further exploration will be necessary.

The lack of an aging deficit for four short sentences poses a puzzle for our findings. No significant age differences were found on any measure of performance for this trial type. As these are null results, only tentative speculations can be made about why these findings might have occurred. One likely possibility involves both short-term and auditory perceptual storage as they relate to specific properties of the sentences. As was seen in the serial position graphs, both young and older adults accessed clauses and recalled more words from those sentences that were presented last, suggesting an effect of recency. It is likely that in these final serial positions, participants were able to retain the auditory memory trace before it decayed. Given the serial position findings, this appeared to occur in all trial conditions. What might separate four short sentences from the remaining three trial types? In the case of four short sentences, the number of chunks presented is smaller than eight short and four random sentences, the chunk size or overall duration is smaller than four long and eight short sentences, and the sentences are also meaningfully organized in comprehensible language, providing an advantage over four random sentences in terms of chunking and encoding. Additionally, as the number of sentences and their respective lengths are small, older adults may be able to bind a majority of the unrelated sentences together without great cognitive cost; this would not be as easy for the remaining trial types, as there are a greater number or length of sentences, and it is likely that the associative deficit would influence chunking and recall in these circumstances. These properties may allow sentences to be more easily encoded and stored in working memory, thus making few demands upon older adults' cognitive resources. In all other cases, more demands are made upon processing and storage into working memory, and this may have greatly impacted older adults' performance for the other three trial types on many of the measures examined.

As was mentioned above, the setup of this study did not allow for exact determination of which of the above factors (chunk limits, length limits, binding ability, among unnamed others) might be involved in task performance. It is probable that all of these factors play a role, but the extent to which each is involved across age groups is unclear. One way to better understand these factors in relation to chunking, as well as their respective degrees of involvement, might be to create mathematical or computational models of chunking stimuli. From the results obtained, a prediction to be tested in further research is that each of the three factors mentioned above operates differentially in young and older adults. This would particularly be the case for binding ability. Young adults will have a greater involvement of binding in a chunking task than older adults, as older adults more typically display deficits in binding and previous research has been unable to replicate an associative deficit in young adults (Naveh-Benjamin et al., 2003, 2007).

One source of limitations in this study's conclusions is due to the use of language materials. One issue is that the sentences used were not controlled for comprising grammatical parts of speech. This may have resulted in some sentences having a composition that would be easier to encode (i.e., noun-verb) than those of other sentences (i.e., noun-verb-direct object). A future study will control not only the parts of speech used in the spoken sentences, but will also utilize critical words. We originally had planned to use critical words as an easier means to measure chunks; these critical words would have been key nouns and verbs that were central to comprehending the sentence, and should be the most likely words to be recalled from a sentence. Because the critical words' parts of speech and their distance from other critical items were not controlled, we decided to consider all words in a sentence instead. The follow-up study will control both of these problems. Using critical items can lead to a better understanding of both gist and verbatim memory of spoken sentences. Older adults have worse verbatim recall than young adults and tend to employ greater reliance upon gist interpretations (Backman &

Nilsson, 1985). If this is the case, we might expect that older adults will recall as many critical items as younger adults, but are much worse in recalling all words within a sentence. These predictions might also vary across the differing sentence conditions. Given the results that were obtained, older adults may not show any significant differences in their gist and verbatim recall relative to young adults on four short sentences, but it is likely that age differences and differences between verbatim and gist recall would be found on the remaining trial types.

Second are the limitations in using language stimuli at all. Because of the numerous connections between phonemes, words, and clauses inherently present in language, it is difficult to determine what might constitute a recalled chunk or where chunk boundaries might be located, despite our attempts to be as accurate as possible with our chunk measures. One possible way to examine this would be to pre-expose participants to a subset of visually presented items (possibly shapes or digits), and then include this subset within a larger set of items. Participants would then either be asked to recall as many items as possible or detect changes in one item in the visual display, similar to the Luck and Vogel (1997) visual change-detection task. Not only will this provide chunking estimates that are not complicated by using verbal stimuli, but this will also allow for a measure of working memory capacity by using Cowan's K (described in Cowan, 2001) or other similar formulas. This paradigm may be able to determine how capacity is changed when pre-exposure to some stimuli permit chunking, and thus, a greater number of items to be stored in working memory. Additionally, we may be able to see whether older adults show impairment in binding pre-exposed items together into a larger unit, suggesting an associative deficit in visual binding as well as confirming previous findings by Cowan et al. (2006).

In agreement with prior studies, older adults encounter greater difficulty in chunking and recalling various spoken sentences relative to young adults, but our results show a more complicated pattern of results. Older adults appear to rely less upon chunking capacity limits, and they show difficulty in forming novel associations between unrelated sentences in a trial condition, but neither of these solely influence the results that were found. In some cases where cognitive demands upon chunking and binding were few, as was the case for four short sentences, older adults formed and recalled as many chunks as young adults; while we are able to make speculations about the cause of these findings, it is not entirely clear why they occurred. More research is needed to determine the circumstances under which each of the influences listed above are involved, as well as how the extent of such involvement might differ across age groups. The findings here present a picture of aging effects on chunking that are incredibly complex; findings from future studies will help to clarify those mechanisms involved.

References

Allen, P. A., & Coyne, A. C. (1988). Age differences in primary organization or processing variability? Part I: An examination of age and primary organization. *Experimental Aging Research, 14,* 143-149.

Allen, P. A., & Coyne, A. C. (1989). Are there age differences in chunking? *Journal of Gerontology, Psychological Sciences, 44,* P181-P183.

Allen, P. A., & Crozier, L. C. (1992). Age and ideal chunk size. *Journal of Gerontology, Psychological Sciences, 47,* P47-P51.

Anderson, J. R., & Matessa, M. (1997). A production system theory of serial memory. *Psychological Review, 104,* 728-748.

Backman, L., & Nilsson, L. G. (1985). Prerequisites for lack of age differences in memory performance. *Experimental Aging Research, 11,* 67-73.

Baddeley, A. (1986). Modularity, mass-action and memory. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 38(A),* 527-533.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning & Verbal Behavior, 14,* 623-629.

Broadbent, D. E. (1975). The magic number seven after fifteen years. In A. Kennedy & A.Wilkes (Eds.) *Studies in long-term memory* (pp. 3-18). New York: Wiley.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4,* 55-81.

Chen, Z., & Cowan, N. (2005). Chunk limits and length limits in immediate recall: A reconciliation. *Journal of Experimental Psychology: Learning, Memory and Cognition, 31,* 1235-1249.

Cohen, G. (1979). Language comprehension in old age. *Cognitive Psychology, 11,* 412-429.

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.) *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62-101). Cambridge, U.K.: Cambridge University Press.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87-185.

Cowan, N. (2005). Working memory capacity limits in a theoretical context. In C. Izawa & N. Ohta (Eds.) *Human learning and memory: Advances in theory and application: The 4th Tsukuba International Conference on Memory* (pp. 155-175). Mahwah, NJ, U.S.: Lawrence Erlbaum Associates Publishers.

Cowan, N., Chen, Z., & Rouder, J. N. (2004). Constant capacity in an immediate serial recall task: A logical sequel to Miller (1956). *Psychological Science, 15,* 634-640.

Cowan, N., Naveh-Benjamin, M., Kilb, A., & Saults, J.S. (2006). Life-span development of visual working memory: When is feature binding difficult? *Developmental Psychology*, *42,* 1089-1102.

Craik, F. I. M. (1983). On the transfer of information from temporary to permanent memory. *Philosophical Transactions of the Royal Society of London, 302,* 341-359.

Craik, F. I. M. (1986). A functional account of age differences in memory. In. F. Klix & H. Hagendorf (Eds.) *Human memory and cognitive capabilities: Mechanisms and performances* (pp. 409-421). New York: Elsevier Science Publishers.

Craik, F. I. M., & Masani, P. A. (1967). Age differences in the temporal integration of language. *British Journal of Psychology, 58,* 291-299.

Craik, F. I. M., Morris, R. G., & Gick, M. L. (1990). Adult age differences in working memory. In G. Vallar & T. Shallice (Eds.) *Neuropsychological impairments of short term memory* (pp. 247-267). New York: Cambridge University Press.

Fallon, M., Kuchinsky, S., & Wingfield, A. (2006). The salience of linguistic clauses in young and older adults' running memory for speech. *Experimental Aging Research, 30,* 359-371.

Glanzer, M., & Razel, M. (1974). The size of the unit in short-term storage. *Journal of Verbal Learning & Verbal Behavior, 13,* 114-131.

Gobet, F., & Clarkson, G. (2004). Chunks in expert memory: Evidence for the magical number four…or is it two? *Memory, 12,* 732-747.

Hay, J. F., & Jacoby, L. L. (1999). Separating habit and recollection in young and older adults: Effects of elaborative processing and distinctiveness. *Psychology & Aging, 14*, 122-134.

Jarvella, R. J. (1970). Effects of syntax on running memory span for connected discourse. *Psychonomic Science, 19,* 235-236.

Jarvella, R. J. (1971). Syntactic processing of connected speech. *Journal of Verbal Learning & Verbal Behavior, 10,* 409-416.

Jennings, J. M., & Jacoby, L. L. (1993). Automatic versus intentional uses of memory: Aging, attention, and control. *Psychology & Aging, 8*, 283-293.

Kausler, D. H. (1994). Long-term episodic memory: Discourse. *Learning and memory in normal aging.* (pp. 276-305). San Diego, CA, U. S.: Academic Press.

Kemper, S. (2006). Language in adulthood. In E. Bialystok & F. I. M. Craik (Eds.) *Lifespan cognition: Mechanisms of change* (pp. 223-238). New York: Oxford

University Press.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390,* 279-281.

Mayberry, M. T., Parmentier, F. B. R., & Jones, D. M. (2002). Grouping of list items reflected in the timing of recall: Implications for models of serial verbal memory. *Journal of Memory and Language, 47,* 360-385.

Miller, G. A. (1956). The magical number seven plus or minus two. Some limits on our capacity for processing information. *Psychological Review, 63*, 81-97.

Mitchell, K. J., Johnson, M. K. , Raye, C. L., Mather, M., & D'Esposito, M. (2000). Aging and reflective processes of working memory: Binding and test load deficits. *Psychology & Aging, 15,* 527-541.

Naveh-Benjamin, M. (2000). Adult age differences in memory performance: Tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory and Cognition, 26,* 1170-1187.

Naveh-Benjamin, M., Cowan, N., Kilb, A., & Chen, Z. (2007). Age-related differences in immediate serial recall: Dissociating chunk formation and capcity. *Memory & Cognition.*

Naveh-Benjamin, M., Craik, F. I. M., Guez, J., & Kreuger, S. (2005). Divided attention in younger and older adults: Effects of strategy and relatedness on memory performance and secondary task costs. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31,* 520-537.

Naveh-Benjamin, M., Guez, J., Kilb, A., & Reedy, S. (2004). The associative deficit of older adults: Further support using face-name associations. *Psychology and Aging, 19*, 541-546.

Naveh-Benjamin, M., Guez, J., & Shulman, S. (2004). Older adults' associative deficit in episodic memory: Assessing the role of decline in attentional resources. *Psychonomic Bulletin & Review, 11*, 1067-1073.

Naveh-Benjamin M., Hussain, Z., & Guez, J. (2003). Adult age differences in episodic memory: Further support for an associative-deficit hypothesis. *Journal of Experimental Psychology:Learning, Memory, and Cognition, 29,* 826-837.

Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics, 2,* 437-442.

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review, 103,* 403-428.

Simon, H.A. (1974). How big is a chunk? *Science, 183*, 482-488.

Stine-Morrow, E. A. L., Gagne, D. D., Morrow, D. G., & DeWall, B. H. (2004). Age differences in rereading. *Memory & Cognition, 32,* 696-710.

Taub, H. A. (1974). Coding for short-term memory as a function of age. *Journal of Genetic Psychology, 125,* 309-314.

Tulving, E., & Patkau, J. E. (1962). Concurrent effects of contextual constraint and word frequency on immediate recall and learning of verbal material. *Canadian Journal of Psychology, 16,* 83-95.

Wingfield, A., & Stine-Morrow, E. A. L. (2000). Language and speech. In F. I. M. Craik & T. A. Salthouse (Eds.) *The handbook of aging and cognition* (2nd ed.) (pp. 359-416). Mahwah, NJ, U.S.: Lawrence Erlbaum Associates Publishers.

Wingfield, A., Wayland, S. C., & Stine, E. A. L. (1992). Adult age differences in the use of prosody for syntactic parsing and recall of spoken sentences. *Journal of Gerontology: Psychological Sciences, 47,* 350-356.

Appendix

Stimuli used for each group and trial type. Stimuli within each trial were presented in the order below; trials were randomized.

Trial Type	Stimuli
	Group 1
4 short, trial 1	Don't scare your brother. We took pictures as a gift. I upset my mother. Answer every question.
4 short, trial 2	Take your paper and pencil. I lied about the money. The teacher enjoys photographs. The screaming bothers me.
8 short, trial 1	Hold the hive still. Add some salted butter. The yellow dog howled. Flashes rattled the sky. Thieves took the painting. My feelings get hurt. We welcomed the family. The worm wiggled below.
8 short, trial 2	The birds circled above. We rested before supper. My temper causes trouble. Dad saw them leave. Lightning struck the building. The witch kicked him. Give the crab flavor. Bees sting when surprised.
4 long, trial 1	Turn the block over and place it by the toy. Grandmother warmed the oven and she baked a fudge cake. I tripped over the rug and I smashed the vase. The drink tumbled down and my shirt was

	stained.
4 long, trial 2	Play in the park but expect rainy weather later. She locked the door and then she walked to school. The woman heard footsteps so she gripped her bag. Our neighbor sells vegetables but he also makes fruit juice.
4 random, trial 1	Go jar she flag. A close football your cheese. The soon help sauce starts. Prepared uncle and cookie.
4 random, trial 2	Girls find cut your. Is my breakfast firewood. Can chicken please finished. Aunt cooked jersey almost.

Group 2

4 short, trial 1	Then she walked to school. My aunt cooked chicken. I smashed the vase. The woman heard footsteps.
4 short, trial 2	She gripped her bag. I tripped over the rug. She prepared a cheese sauce. She locked the door.
8 short, trial 1	Please find your jersey. The drink tumbled down. Place it by the toy. Play in the park. Go and help your uncle. Our neighbor sells vegetables. She baked a fudge cake. Breakfast is almost finished.
8 short, trial 2	Close the cookie jar. Grandmother warmed the oven. He also makes fruit juice. Girls can cut firewood. Expect rainy weather later.

	Turn the block over. My shirt was stained. Flag football starts soon.
4 long, trial 1	My temper causes trouble when my feelings get hurt. We welcomed the family and we rested before supper. Give the crab flavor and add some salted butter. The birds circled above as the worm wiggled below.
4 long, trial 2	Don’t scare your brother because the screaming bothers me. The yellow dog howled as the witch kicked him. I upset my mother when I lied about the money. Thieves took the painting and Dad saw them leave.
4 random, trial 1	Took when pencil and surprised. Gift building a as. The struck pictures your sting. Lightning paper we bees take.
4 random, trial 2	The hold sky hive. Answer the still flashes. Photograph teacher every. Enjoys the rattled question.

Table 1: Reassignment of Group 1 stimulus materials to form Group 2 stimuli.

Group 1		Group 2
4 Long	divided among	4 Short, 8 Short
4 Random	divided among	4 Short, 8 Short
4 Short	divided among	4 Random, 4 Long
8 Short	divided among	4 Random, 4 Long

See Appendix for details

Table 2: Means and standard errors for total words, adopted chunks, chunk size, clause access, and clause completion across age groups and trial conditions. Standard errors are in parentheses.

Age Group	**Condition**	**Total Words**	**Adopted Chunks**	**Chunk Size**	**Clause Access**	**Clause Completion**
Young adults	*Four Long*	15.97 (1.04)	2.75 (.167)	6.07 (.387)	4.64 (.263)	.709 (.024)
	Four Random	5.66 (.301)	2.20 (.147)	2.80 (.211)	2.12 (.113)	.652 (.034)
	Four Short	10.68 (.438)	2.87 (.157)	3.96 (.286)	3.10 (.099)	.809 (.020)
	Eight Short	10.93 (.658)	3.20 (.190)	3.50 (.187)	3.22 (.167)	.809 (.032)
Older adults	*Four Long*	12.06 (.785)	2.37 (.151)	5.31 (.33)	3.52 (.191)	.727 (.025)
	Four Random	3.81 (.434)	1.64 (.186)	2.48 (.223)	1.50 (.164)	.667 (.046)
	Four Short	10.35 (.638)	2.56 (.176)	4.26 (.273)	2.93 (.154)	.811 (.021)
	Eight Short	8.41 (.775)	2.56 (.28)	3.45 (.174)	2.45 (.255)	.821 (.036)

Figure Captions:

Figure 1. Hypothetical data based on expected outcomes in terms of a chunk-limited working memory capacity, using mean words recalled as an example measure.

Figure 2. Hypothetical data based on expected outcomes if a length or time-limited working memory capacity solely influences formation and recall of chunks.

Figure 3. Predictions based on a combination of chunk capacity limits, length limits, and the associative deficit hypothesis.

Figure 4. Mean words recalled by condition across age groups. Error bars are standard errors of the mean.

Figure 5. Mean adopted chunks formed by condition across age groups. Error bars are standard errors of the mean.

Figure 6. Mean chunk size by condition across age groups, with chunk size measured as the number of words per adopted chunk. Error bars are standard errors of the mean.

Figure 7. Mean number of clauses accessed by condition across age groups. Error bars are standard errors of the mean.

Figure 8. Mean proportion of words recalled conditional on clauses that were accessed. Error bars are standard errors of the mean.

Figure 9. Mean access per clause for each sentence condition. Dashed vertical lines shown on the graph for four long sentences denote sentence boundaries. Error bars are standard errors of the mean.

Figure 10. Mean words recalled per clause for each sentence condition. Dashed vertical lines shown on the graph for four long sentences denote sentence boundaries. Error bars are standard errors of the mean.

Figure 1:

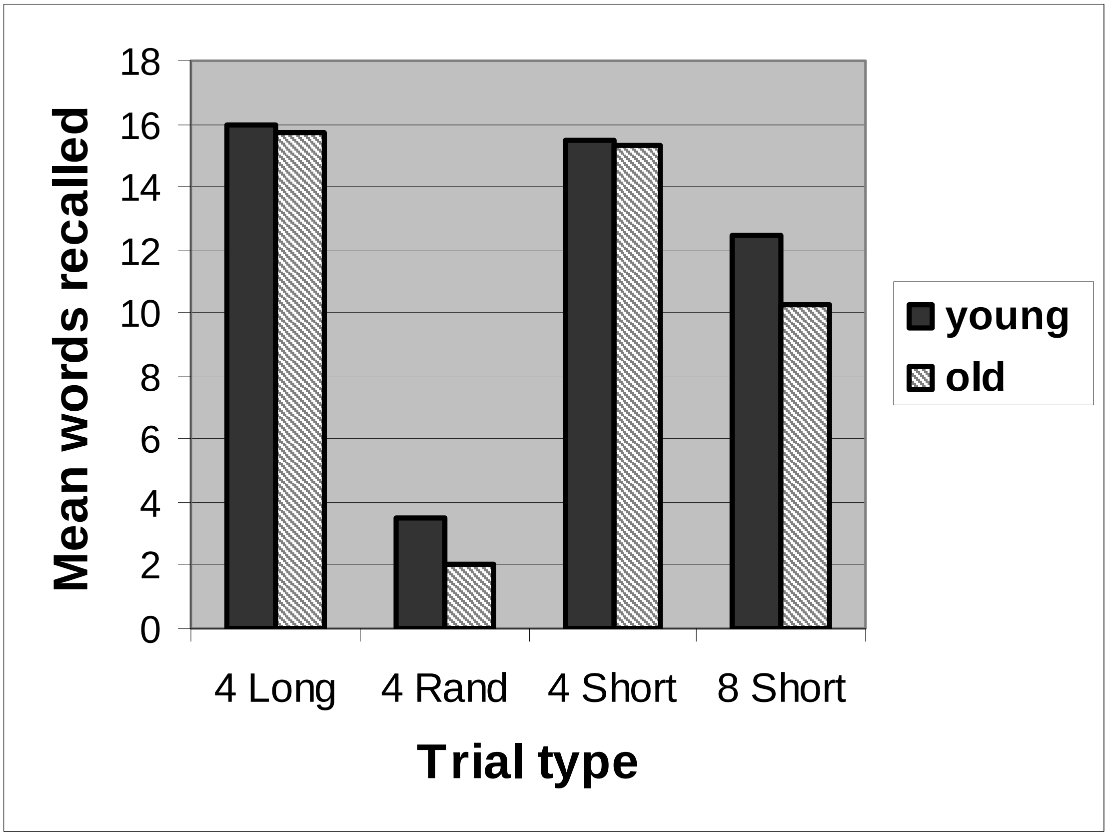

18
16
14
12
10
8
6
4
2
0
Mean words recalled
4 Long
4 Rand
4 Short
8 Short
Trial type
young
old

Figure 2:


18
16
14
12
10
8
6
4
2
0
Mean words recalled
4 Long
4 Rand
4 Short
8 Short
Trial type
young
old

Figure 3:


Mean words recalled
18
16
14
12
10
8
6
4
2
0
4 Long
4 Rand
4 Short
8 Short
Trial type
young
old

Figure 4:


18
16
14
12
10
8
6
4
2
0
Mean words recalled
4 Long
4 Rand
4 Short
8 Short
Trial Type
young
old

Figure 5:

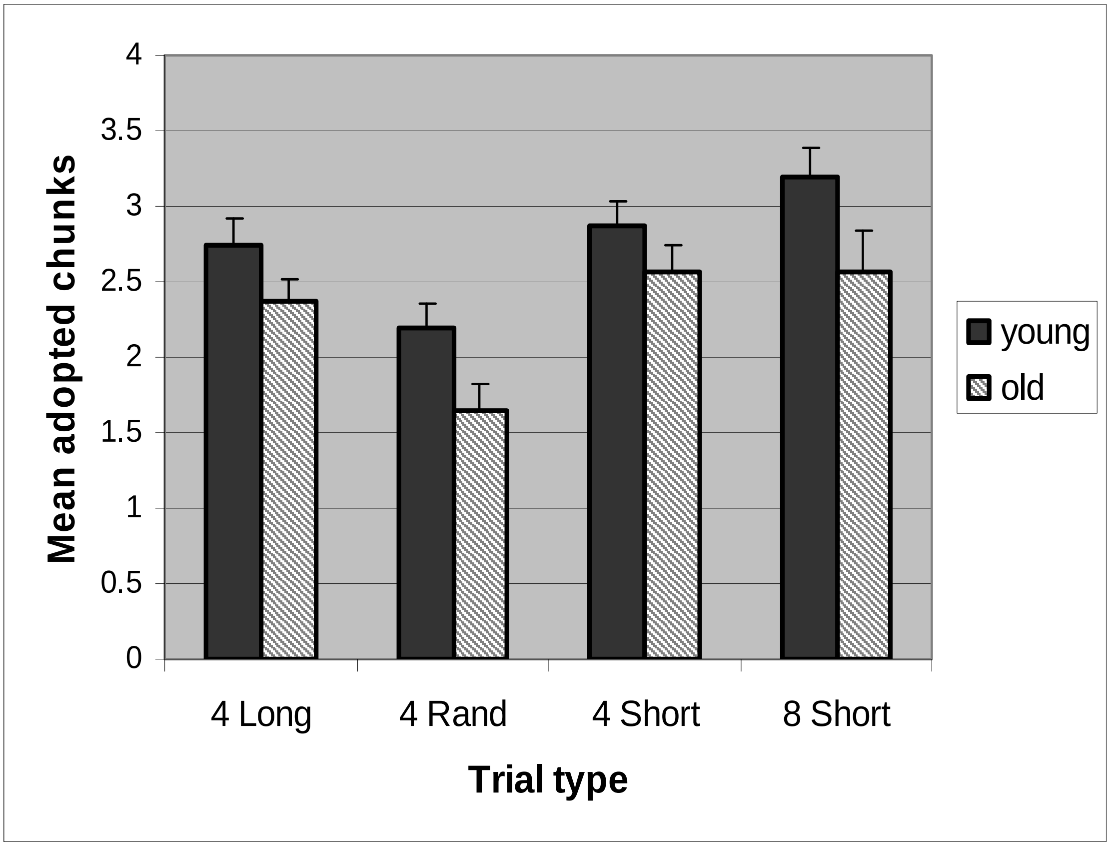

4
3.5
3
2.5
2
1.5
1
0.5
0
Mean adopted chunks
4 Long
4 Rand
4 Short
8 Short
Trial type
young
old

Figure 6:

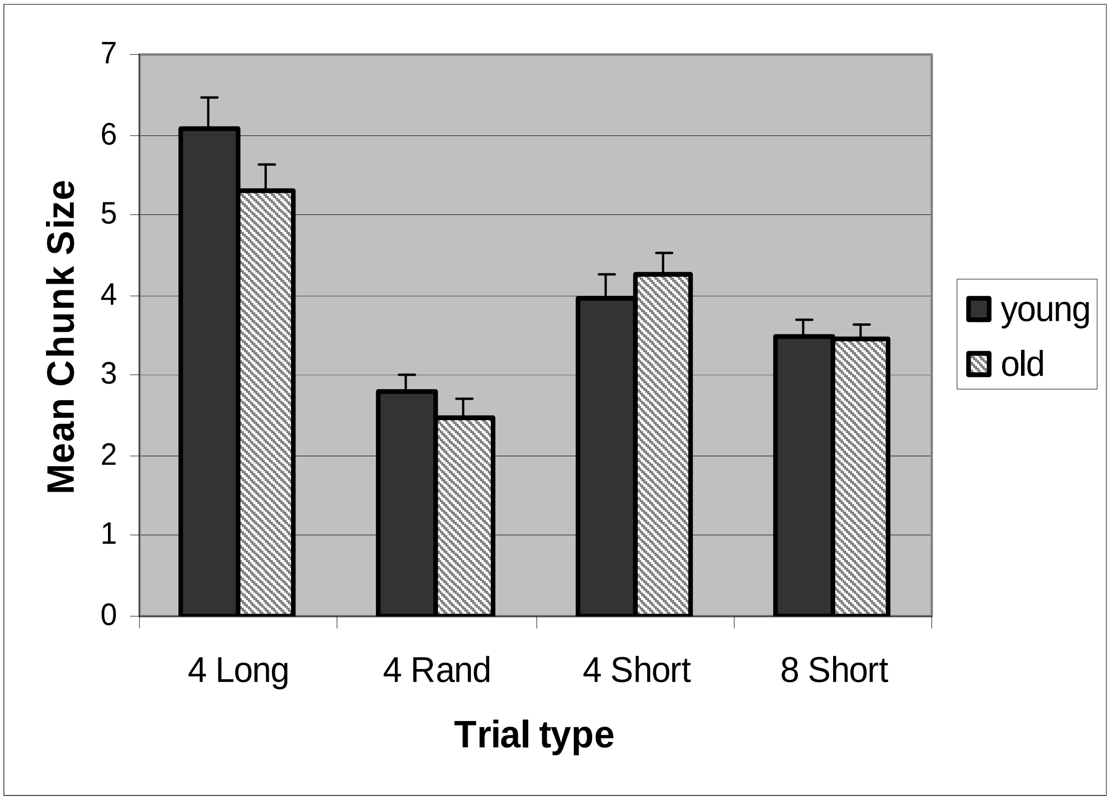

7
6
5
4
3
2
1
0
Mean Chunk Size
young
old
4 Long
4 Rand
4 Short
8 Short
Trial type

Figure 7:

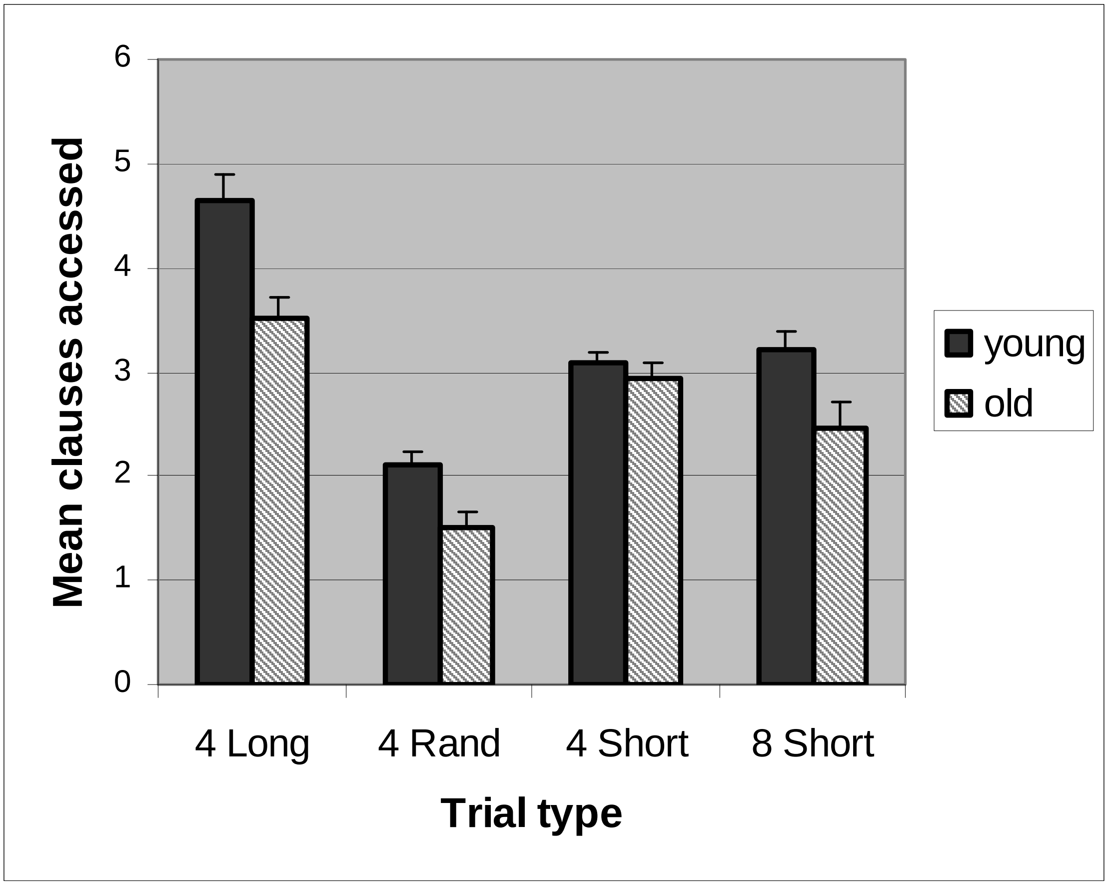

6
5
4
3
2
1
0
Mean clauses accessed
4 Long
4 Rand
4 Short
8 Short
Trial type
young
old

Figure 8:

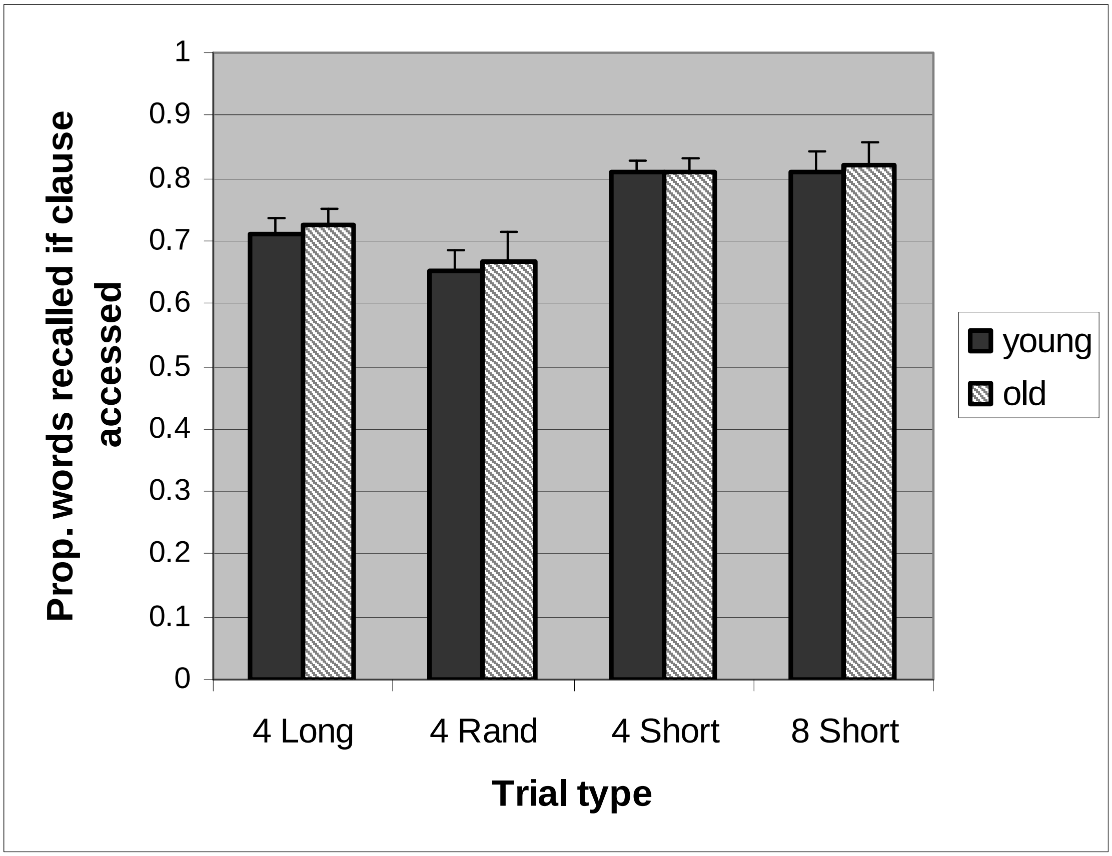

1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Prop. words recalled if clause accessed
young
old
4 Long
4 Rand
4 Short
8 Short
Trial type

Figure 9:

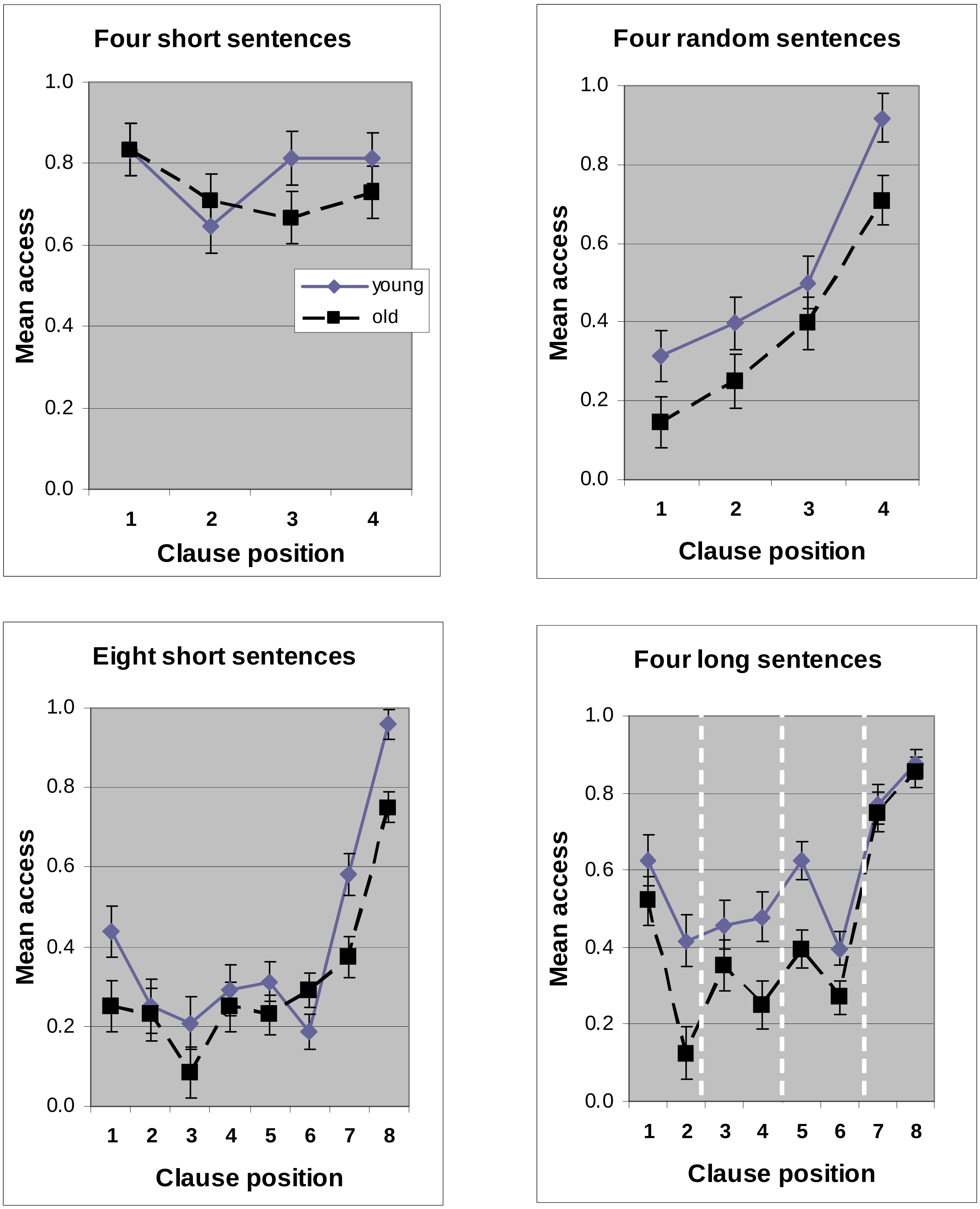
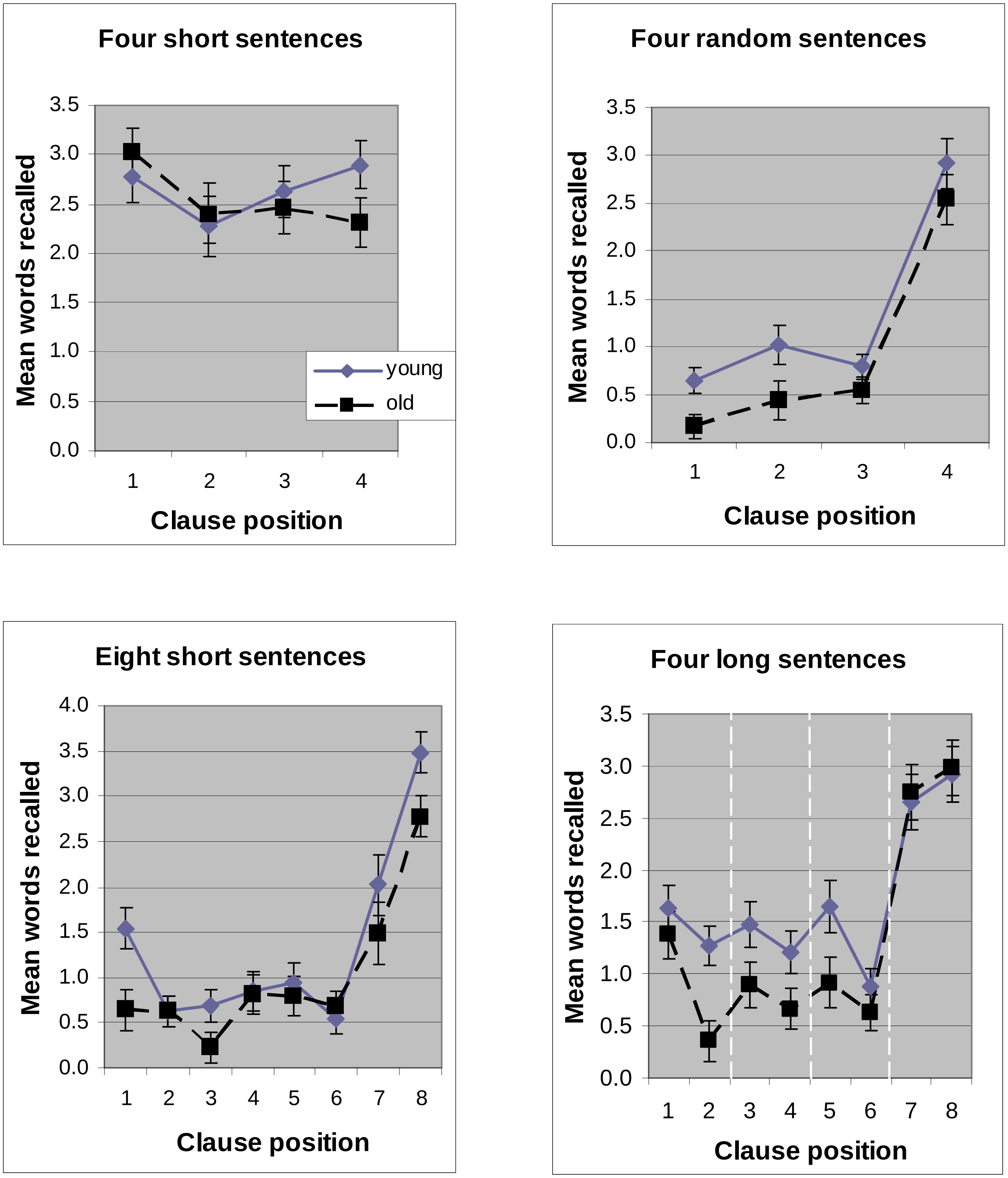

Four short sentences
Mean access
1.0
0.8
0.6
0.4
0.2
0.0
1
2
3
4
Clause position
young
old
Four random sentences
Mean access
1.0
0.8
0.6
0.4
0.2
0.0
1
2
3
4
Clause position
Eight short sentences
Mean access
1.0
0.8
0.6
0.4
0.2
0.0
1
2
3
4
5
6
7
8
Clause position
Four long sentences
Mean access
1.0
0.8
0.6
0.4
0.2
0.0
1
2
3
4
5
6
7
8
Clause position

Figure 10:


Four short sentences
Mean words recalled
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
1
2
3
4
Clause position
young
old
Four random sentences
Mean words recalled
Clause position
Eight short sentences
Mean words recalled
4.0
1
2
3
4
5
6
7
8
Clause position
Four long sentences
Mean words recalled
Clause position