Adult Age Differences in the Benefit of Syntactic and Semantic Constraints for Sentence Processing

Caroline Beese
Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany, and Max Planck Institute for Human Development, Berlin, Germany

Markus Werkle-Bergner
Max Planck Institute for Human Development, Berlin, Germany

Ulman Lindenberger
Max Planck Institute for Human Development, Berlin, Germany, and Max Planck UCL Centre for Computational Psychiatry and Ageing Research, Berlin, Germany

Angela D. Friederici and Lars Meyer
Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

Verbal working memory-intensive sentence processing declines with age. This might reflect older adults’ difficulties with reducing the memory load by grouping single words into multiword chunks. Here we used a serial order task emphasizing syntactic and semantic relations. We evaluated the extent to which older compared with younger adults may differentially use linguistic constraints during sentence processing to cope with verbal working memory limitations. Probing syntactic–semantic interactions, age differences were hypothesized to be confined to the use of syntactic constraints and to be accompanied by an increased reliance on semantic information. Two experiments varying in verbal working memory demands were conducted: the sequence length was increased from eight items in Experiment 1 to 11 items in Experiment 2. We found the use of syntactic constraints to be compromised with aging, while the benefit of semantic information for sentence processing was comparable across age groups. Hence, we suggest that semantic information processing may become relatively more important for successful sentence processing with advancing adult age, possibly inducing a syntactic-to-semantic-processing strategy shift.

Keywords: aging, chunking, semantics, syntax, working memory

It is widely debated whether language processing declines with age (for review, see Shafto & Tyler, 2014). While aging has been argued to spare the processing of short or simple sentences (e.g., Gilchrist, Cowan, & Navh-Benjamin, 2008; Wingfield, Poon, Lombardi, & Lowe, 1985), the processing of long or complex sentences deteriorates with age (e.g., Feier & Gerstman, 1980; Kemper, 1987; Obler, Fein, Nicholas, & Albert, 1991). This may be related to an age-related decline in verbal working memory (vWM; e.g., Beese, Meyer, Vassileiou, & Friederici, 2017; Bopp & Verhaeghen, 2005; Salhouse, 1994) because both long and complex sentences tax vWM (Gibson, 2000; Just & Carpenter, 1992; Meyer, Obleser, Kiebel, & Friederici, 2012). The present study investigated whether older adults exploit specific linguistic constraints, that is syntactic regularities and semantic relations, to different extents to counteract their reduced vWM capacities when sentence processing becomes difficult.

Syntactic and semantic constraints on sentence processing have been shown to be employed differentially across the life span (e.g., Friederici, Schriefers, & Lindenberger, 1998). Comparing syntactic and semantic processing directly, it was found that syntactic, but not semantic, processing was deficient in older, relative to younger adults (Friederici et al., 1998). Other studies investigating either syntactic or semantic processing in elderly people support this finding: Syntactic processing speed and accuracy are lower in older compared with younger adults when processing demands are high; for example, when processing syntactically long, ambiguous, or complex sentences (e.g., Kemper, 1986; Kemper, Crow, &
Kemtes, 2004; Obler et al., 1991, respectively). In comparison, lexical-semantic processing speed and accuracy do not appear to differ between younger and older adults (e.g., Burke, White, & Diaz, 1987; Gunter, Jackson, & Mulder, 1998; Laver, 2009). In addition, the results of numerous studies document adult age differences in online sentence processing heuristics during self-paced reading (Radvansky, 1999; Radvansky, Curiel, Zwaan, & Copeland, 2001; Radvansky & Dijkstra, 2007; Stine, 1990; Stine-Morrow, Loveless, & Soederberg, 1996; Stine-Morrow, Noh, & Shake, 2010). Specifically, older adults allocate fewer resources than younger adults to the decoding of orthographic or syntactic features (i.e., the surface form) or to the conceptual integration of words into coherent syntactic constituents conveying propositional content (i.e., the textbase). Rather, older adults focus on the creation of the mental representation of the events described (i.e., the situation model). Hence, older adults may rely more on semantic than syntactic constraints to successfully process sentences.

Insufficient resources, in particular verbal working memory limitations as a domain-general processing resource (Gibson, 2000; Just & Carpenter, 1992; Meyer et al., 2012), may impair older adults’ engagement in enriched sentence processing (e.g., Christianson, Williams, Zacks, & Ferreira, 2006). Generally, the verbatim surface form and textbase information (i.e., syntactic regularities and propositional content, respectively) are remembered for shorter intervals than the situation model of sentences (i.e., the overall event described; Kintsch, Welsch, Schmalhofer, & Zimny, 1990; Sachs, 1967). Particularly, in older adults, representations of surface form or textbase information fade away even more rapidly (e.g., Radvansky et al., 2001). Hence, it seems plausible to assume that advancing adult age is accompanied by a shift in processing heuristics. According to this proposal, older adults allocate fewer resources to an enriched decoding of surface form or textbase information and more resources to the situation model than do younger adults (e.g., Stine-Morrow et al., 1996). This processing heuristic may help older adults to reduce demands on resources that would otherwise be unavailable if already consumed by these lower level processes.

Syntactic and semantic constraints can be helpful to reduce vWM costs during sentence processing. That is, discrete words can be grouped into multiword phrases based on syntactic regularities and semantic relations (e.g., Bonhage, Meyer, Gruber, Friederici, & Mueller, 2017; Epstein, 1962). This grouping, henceforth chunking, increases the amount of information stored in working memory (Gilchrist, 2015; Gilchrist & Cowan, 2012; Gobet et al., 2001; Gobet, Lane, & Lloyd-Kelly, 2015; Miller, 1956). On the one hand, words can be grouped together on the basis of syntactic regularities (e.g., word categories, word order); for example, the function word “the” with the content word “boat” into the multiword phrase “the boat” (e.g., Bonhage et al., 2017; Meyer, Henry, Gaston, Schmuck, & Friederici, 2017; Schell, Zaccarella, & Friederici, 2017). When words are grouped together by syntactic regularities, they are remembered better than words that cannot be grouped together in this way (the so-called sentence superiority effect, e.g., Baddeley, Hitch, & Allen, 2009; Bonhage, Fiebach, Bahlmann, & Mueller, 2014; Bonhage et al., 2017; Cattell, 1886; Snell & Grainger, 2017). From now on, we call this grouping process syntactic chunking. The term denotes the abstract binding of words’ syntactic categories, which has been argued to occur rather automatically, unconsciously, and without any strategic efforts (Baddeley et al., 2009; Freudenthal, Pine, Aguado-Orea, & Gobet, 2007; Gobet, Lloyd-Kelly, & Lane, 2016). In the previous example, the binding of “the” and “boat” is based on syntactic not semantic information, creating a two-word phrase without building up a semantic representation that would go beyond the word-level meaning. Similarly, multiple content words like “sleepy” and “boat” may also be grouped together based on syntactic information. However, this can result in a semantically anomalous phrase, like “sleepy boat,” when the content words are not semantically related. But when words can indeed be grouped together based on semantic information (e.g., semantic relatedness), then enriched semantic representations are formed. For example, the content words “wooden” and “boat” may be grouped together into one joint representation of a “wooden boat” creating an enriched semantic representation of “boat” (Schell et al., 2017). This is conventionally known as conceptual chunking (Bierwisch & Schreuder, 1992; Miller, 1956).

Age differences in the chunking of word sequences have been found in studies examining either online, real-time processing (for an overview, see Stine-Morrow & Payne, 2016) or poststimulus, offline effects (Gilchrist et al., 2008); yet, the online and offline results differ. On the one hand, online studies suggest that older adults generate smaller chunks to accommodate their vWM capacity limitations, reflected in a greater allocation of processing resources to intrasentence compared with intersentence boundaries (e.g., Stine, 1990; Stine-Morrow et al., 2010). On the other hand, offline studies suggest that older adults remember fewer, but not smaller chunks than younger adults. For instance, Gilchrist et al. (2008) manipulated both memory load (i.e., number of unrelated sentences) and chunk size (i.e., number of clauses), reporting age differences during verbatim recall with respect to the former but not to the latter. There are several relevant differences between previous online and offline studies. First, online and offline studies assessed chunking differently: Online studies inferred chunking on the basis of reading or gaze time measurements, while offline studies assessed chunking by enhancing or reducing opportunities for chunking in the experimental material. Second, online studies dissociated syntactic and semantic information processing via statistical means, rather than via experimental manipulations; the offline studies did not dissociate the contribution of these distinct linguistic information types on chunking at all. However, no study to date has assessed age differences in the extent to which different linguistic information types are used for chunking, neither online nor offline. To clarify this open question, our study aimed at dissociating the extent to which older adults differentially use syntactic and semantic constraints for chunking during sentence processing to circumvent their vWM limitations, using an offline task.

Syntactic chunking has been shown to occur independently of semantic information processing in younger adults (Epstein, 1962); however, this may change with increasing syntactic processing deficits across the life span. Here, we manipulate the availability of syntactic and semantic constraints for chunking during sentence processing in younger and older adults. We employed sentences that pose lower (Experiment 1) or higher (Experiment 2) vWM demands through varying the sequence length. Syntactic chunking was assessed through the classical comparison of multiword syntactic phrases versus random word lists (Baddeley et al., 2009). To assess semantic information processing, we
adapted the classical dissociation of meaningful versus meaningless words by comparing real words to pseudowords (e.g., Cattell, 1886; Epstein, 1962; Reicher, 1969). We used an offline serial order task to assess processing success for verbal materials that vary in the availability of syntactic and semantic constraints. The presence of those linguistic constraints facilitated or hampered chunking and therefore the accommodation of vWM limitations. We hypothesized that older adults do not use syntactic constraints as much as younger adults during the processing of longer sentences, because vWM-intensive sentence processing is known to deteriorate with age (e.g., Kemper, 1986; Kemper et al., 2004; Obler et al., 1991). Furthermore, because of this disadvantage in syntactic chunking, we expected a stronger reliance on semantic constraints in older participants. This is in line with previous studies suggesting a syntactic-to-semantic processing strategy shift in the elderly (e.g., Amichetti, White, & Wingfield, 2016). Systematically varying the availability of syntactic and semantic constraints during sentence processing, the results showed that the advantage of remembering the temporal order of words in sentences over lists was reduced in older compared with younger adults, in particular when sentences were longer, and thus highly vWM-demanding (Experiment 2). This observation indicates that syntactic constraints may not be used as much by older compared with younger adults. While the presence of syntactic constraints was more beneficial for younger than older adults, the benefit of the presence of semantic constraints was comparable across age groups.

**Experiment 1**

Previous literature suggests that older adults have difficulties processing vWM-intensive sentences (e.g., Beese et al., 2017; Kemper, 1986, 1987; Wingfield et al., 1985). The extent to which the presence of syntactic regularities benefits vWM-intensive sentence processing, by enabling syntactic chunking and thereby reducing the vWM load, can be indexed through the sentence superiority effect (Baddeley et al., 2009). Age differences in the sentence superiority effect have been found more pronounced in eight-word sequences (Wingfield et al., 1985). Specifically, it has been shown that older adults repeat five- and eight-word sentences as accurately as younger adults. However, older but not younger adults’ repetition accuracy is poorer for eight-word but not five-word lists. Therefore, in Experiment 1, meaningful and meaningless lists and sentences of eight words were compared. We aimed at dissociating the contribution of syntactic from semantic constraints to age differences in sentence processing.

**Method**

**Participants.** Fifty-six healthy right-handed native speakers of German, divided into two equally sized age groups (younger vs. older adults), participated for a reimbursement of 9 € per hour. After outlier removal (see Experiment 1 – Statistical analysis), the data of 27 younger adults (13 men, age range = 20 – 31 years, mean age = 25.67 years, SD age = 2.56 years) and 26 older adults (12 men, age range = 62 – 71 years, mean age = 66.35 years, SD age = 2.40 years) were analyzed. All participants had normal or corrected-to-normal vision and did not suffer from any neurological disorders according to self-report. The age groups were matched for their level of education (i.e., they had at least a high-school diploma or an equivalent; for more details, see Table 1). Prior to the experiment, all participants gave written informed consent. The experiment was approved by the ethics committee of the University of Leipzig. The experiment was conducted in accordance to the Declaration of Helsinki.

**Verbal working memory measures.** To assess potential associations of age differences in vWM with sentence processing (e.g., Bopp & Verhaeghen, 2005; Lewis, Vasishth, & Van Dyke, 2006; Salthouse, 1994), we acquired three additional standardized measures. Two of those measured particularly the storage component of vWM, or otherwise short-term memory (STM) capacity (i.e., the digit span task forward; Aster, Neubauer, & Horn, 2006; and the nonword repetition span task; Welte, 1981). One further test assessed both the storage as well as the manipulation component (i.e., the digit span task backward; Aster et al., 2006). Given that the three tests of vWM showed a moderate degree of common variance (α = .69), we combined the test scores into one composite score (see Experiment 1 – Statistical analysis). The test results are reported in Table 1. For 18 participants (12 younger, 6 older adults), these tests were acquired for the current study in this order; for all other participants, the test results were available from a previous study (mean delay = 13 months; SD delay = 1 month; Beese et al., 2017).

**Stimuli.** The study design (Table 2) followed a 2 x 2 design with the factors sentence structure (sentences vs. word lists) and meaning (real words vs. pseudowords).

<table>
<thead>
<tr>
<th>Measure</th>
<th>Young (n = 27)</th>
<th>Old (n = 26)</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>25.67 (2.56)</td>
<td>66.35 (2.40)</td>
<td>−.67</td>
<td>.50</td>
</tr>
<tr>
<td>Education</td>
<td>17.56 (1.99)</td>
<td>17.92 (1.98)</td>
<td>.08</td>
<td>.01</td>
</tr>
<tr>
<td>Digit span forward</td>
<td>10.81 (2.29)</td>
<td>9.85 (1.59)</td>
<td>1.79</td>
<td>.01</td>
</tr>
<tr>
<td>Digit span backward</td>
<td>8.70 (2.27)</td>
<td>6.15 (1.54)</td>
<td>4.80</td>
<td>.01</td>
</tr>
<tr>
<td>Repetition span</td>
<td>27.93 (3.00)</td>
<td>25.62 (3.59)</td>
<td>2.54</td>
<td>.01</td>
</tr>
<tr>
<td>Composite score</td>
<td>.37 (0.72)</td>
<td>−.38 (0.66)</td>
<td>3.94</td>
<td>2.50 \times 10^{-4}</td>
</tr>
</tbody>
</table>

Table 1

**Demographics and Individual Differences: Summary of Experiment 1 and 2**
Table 2

Experiment 1, Example of Stimulus Material: Meaningful Sentences (Structure+, Meaning+), Meaningful Word Lists (Structure–, Meaning+), Meaningless Sentences (Structure+ Meaning–), and Meaningless Word Lists (Structure–, Meaning–)

<table>
<thead>
<tr>
<th>Meaning</th>
<th>Sentences</th>
<th>Lists</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real words</td>
<td>der Opa verdarb die Suppe mit dem Salz</td>
<td>der dem Suppe mit Opa Salz die verdarb</td>
</tr>
<tr>
<td>Pseudowords</td>
<td>der Apo verworb die Junne mit dem Sohr</td>
<td>der dem Junne mit Apo Sohr die verworb</td>
</tr>
</tbody>
</table>

Stimuli consisted of eight items; that is, words or pseudowords. All sentence conditions had a consistent syntactic structure (e.g., [der Opa] [verdarb [die Suppe] [mit dem Salz]]; the granddad ruined the soup with the salt). The eight items spanned three phrases connected through the verb. Meaningless sentence conditions were created by keeping function words in place while replacing content words with pseudowords (e.g., [der Apo] [verworb [die Junne] [mit dem Sohr]]; the apo verworbed the junne with the sohr). Pseudowords were generated based on the real content words using Wuggy (Keuleers & Brysbaert, 2010), matching the original words in syllable length and syllable transition frequency. Each content word and pseudoword was unique across the whole stimulus set. List conditions were created by permuting all items of each sentence, keeping the lexical information equal across real word conditions (e.g., real word list: der dem Suppe mit Opa Salz; die verdarb; the soup with granddad salt the ruined). Permutation patterns were identical across the real word and pseudoword conditions (e.g., pseudoword list: der dem Junne mit Apo Sohr die verworb; the junne with apo sohr the verworbed). To avoid syntactic phrase formation, a number of permutation patterns were excluded (verb-noun, verb-determiner, noun-verb, and verb-preposition; cf. Brennan et al., 2012). This reduced opportunities for syntactic chunking in list conditions. However, because of the flexible word order in German, less common partial phrases were inevitable (e.g., die verdarb could, in another context, also be part of a relative clause). Nevertheless, list conditions constituted a stark contrast to the sentence conditions in terms of the availability of syntactic regularities used for syntactic chunking. To avoid encoding differences introduced by the stimulus onset across conditions, the first item was always a lower-case determiner.

The permutation of real word sentences into real word lists aimed at dissolving information provided by the syntactic structure, while preserving the lexical meaning. Yet, the permutation could have induced a confound to our manipulation of lexical meaning. That is because each incoming word adds to the amount of lexical-semantic information of the current context, in both real word sentences and lists. We controlled for any differences in the amount of lexical-semantic information provided by the context, between real word sentences and lists, by employing the lexical surprisal metric (e.g., Hale, 2001), a computational measure of a word’s unexpectedness given a current context. Specifically, differences in surprisal between real word sentences and lists were minimized to an absolute z-transformed surprisal value smaller than 2. To this end, a freely available lexicalized top-down left-corner parser (Roark, Bachrach, Cardenas, & Pallier, 2009), trained on the TIGER Treebank (Brants et al., 2004), was employed. To control for effects of animacy (Bonin, Gelin, & Bugaiska, 2014), half of the subjects were animate, half inanimate. To account for possible differences in lexical access demands, word length (i.e., syllable count = 1 – 3 syllables per word; Spoehr & Smith, 1973) and word frequency (i.e., frequency class = 8 – 15; Goldhahn, Eckart, & Quasthoff, 2012; Halgren et al., 2002) were constrained in range across items to control for variability. According to these criteria, 96 sequences in the four conditions were created, resulting in a total pool of 384 stimuli.

Participants’ ability to use syntactic chunking and semantic information to facilitate vWM-intensive sentence processing was evaluated via a serial order task (Bonhage et al., 2014, 2017): After each stimulus, participants were asked to judge whether the order of two items from the last stimulus matched the order of presentation. For example, after the stimulus der Opa verdarb die Suppe mit dem Salz, the participants may have been asked: Did the word Opa come before the word mit? (abbreviated on screen as Opa <<< mit). Probing the serial order position emphasized the decoding of syntactic regularities and semantic relations during sentence processing. A total of eight different questions were created for each stimulus, querying any pair of items that were divided by three other items. The prompted words were always selected from different phrases in sentence conditions, assuring similar retrieval difficulty compared with list conditions (Johnson, 1978). For half of those questions, the correct answer was yes, for the other half no. The assignment of questions resulted in a total of 3,072 stimulus–question combinations.

The total set of stimulus–question combinations was distributed across 16 stimulus lists of 192 stimuli each; thus, each stimulus list contained always two conditions (one real word condition, one pseudoword condition) out of one set of the four conditions. The question assignment was counterbalanced within lists across sequences and across sequences across lists. In addition, the sequences of each stimulus list were pseudorandomized such that no condition, question, answer type or animate subject occurred consecutively more than four to five times and such that the two conditions of the same stimulus were always separated by at least 20 other sequences.

Procedure. Before the experiment, participants were informed about the four experimental conditions and the task. Stimuli were presented visually, item by item, for 600 ms per item, in white, nonserif letters (Arial, font size 30 pt) in the center of the screen (Sony, Trinitron, 17 in., 60 Hz refresh rate) on a gray background. The stimulus presentation was followed by a fixation
cross of 2,500 ms, followed by a question probing serial order position. Participants had a time limit of 8 s to answer the question with either yes or no by pressing either of two buttons on a button box. Button assignment was counterbalanced across participants. After the response, participants were provided with visual feedback (happy or sad emotion icon) for 400 ms. An intertrial interval of 2,500 ms followed each trial. A practice session of eight trials familiarized participants with the procedure. At five equally distributed points across the experiment, participants could take a break of a maximum of 4 min.

Statistical analysis. For statistical analyses, d-prime (d’) scores were calculated from participants’ yes and no responses to the yes–correct and no–correct questions; after reaction time (RT) outliers (i.e., RTs outside the range of the median RT +/- 2.5 median absolute deviations (MADs; Leys, Ley, Klein, Bernard, & Licata, 2013), and missing responses were removed (both constituted on average 1.68% per participant, SD = 1.66%). d’ scores account for participants’ inherent response bias (Macmillan & Creelman, 2005) and are computed by subtracting the z-transformed false-alarm rate (FA; i.e., yes responses to no–correct questions) from the z-transformed hit rate (H; i.e., yes–responses to yes–correct questions). H and FA of 0 were corrected by 1/(N + 1), while H and FA of 1 were corrected by (N-1)/N, whereby N is the number of trials (Macmillan & Creelman, 2005). Within each age group, participants who performed below chance (i.e., d’ < 0, two older adults), as well as those which constituted group outliers (i.e., d’ > 2 SD, one younger adult and the same two older below chance performers) were excluded from further analyses (in total three participants).

d’ scores were subjected to a three-way analysis of variance (ANOVA), including the within-participants factors sentence structure (sentences vs. word lists) and meaning (real words vs. pseudowords) and the between-participants factor age group (younger vs. older adults). To assess a potential influence of age differences in vWM capacity (Bopp & Verhaeghen, 2005) on d’ scores, the three additionally acquired memory measures were first z-transformed across both age groups simultaneously and then averaged into one composite score in order to be later included as a between-subjects covariate into an analysis of covariance (ANCOVA). Any interaction effects were broken down with simple effects analysis, controlling for multiple comparisons with Bonferroni correction.

Post hoc, we dissociated the impact of the availability of syntactic and semantic constraints for chunking from mere effects of syntactic and lexical predictability; in principle, participants’ memory retrieval success could depend on the syntactic or lexical predictability of a word to occur within a prior sentence. To assess this potential confound, we computed the lexical-semantic and syntactic predictabilities of the word that was to be retrieved for a given trial, using computational-linguistic methodology (for more details on the calculation of these measures, see Meyer & Gumbert, 2018; Roark et al., 2009). For example, for the question Did the word Opa come before the word mit?, predictabilities of the word Opa (granddad) of the prior sentence der Opa verdarb die Suppe mit dem Salz (the granddad ruined the soup with the salt) were computed. This procedure was repeated for all trials of the structured and meaningful condition. For each participant, both accuracy and surprisal values were then averaged across trials within word position. Across word positions, within participants, we then calculated correlations between accuracy and surprisal values, reasoning that if retrieval demands were confounded with predictability, predictability should significantly correlate with accuracy. Across participants, we corrected the test statistics for multiple comparisons using the False Discovery Rate (FDR) procedure.

Results

The three-way ANOVA (sentence structure × meaning × age group) on the d’ values showed three main effects: sentence structure, $F(1, 51) = 182.95$, $p = 2.20 \times 10^{-16}$, $\eta^2 = .39$; meaning, $F(1, 51) = 57.79$, $p = 6.12 \times 10^{-10}$, $\eta^2 = .60$; and age group, $F(1, 51) = 6.88$, $p = 1.15 \times 10^{-2}$, $\eta^2 = .09$. Figure 1 shows an overview of the single subject and group mean (young vs. old) d’ values for each condition (real word sentences, pseudoword sentences, real word lists, and pseudoword lists). Table 3 shows an overview of all response types (hit, miss, false alarm, and correct rejection rates) for Experiment 1 and 2 across age groups.

More precisely, performance for sentences ($M = 3.20$, $SD = 1.42$) was higher than for lists ($M = 1.54$, $SD = 0.79$). Performance for meaningful stimuli ($M = 2.65$, $SD = 1.51$) was higher than for meaningless stimuli ($M = 2.10$, $SD = 1.28$). Younger adults performed better ($M = 2.69$, $SD = 1.42$) than older adults ($M = 2.04$, $SD = 1.35$). Furthermore, an interaction effect was found between structure and meaning, $F(1, 51) = 17.80$, $p = 1.00 \times 10^{-4}$, $\eta^2 = .02$. Simple effects analysis (Bonferroni-corrected) showed that while performance was significantly higher for meaningful sentences ($M = 3.64$, $SD = 1.40$) than meaningless sentences ($M = 2.77$, $SD = 1.32$); $t(102) = -8.33$, $p = 1.55 \times 10^{-16}$.

Figure 1. Experiment 1, trending age differences in syntactic (sentences vs. lists) but not semantic (real words vs. pseudowords) processing.
10^-12, this performance difference was not present for meaningful lists (M = 1.66, SD = 0.82) compared with meaningless lists (M = 1.42, SD = 0.75); t(102) = -2.30, p < .09. Moreover, simple effects analysis showed a performance advantage for sentences over word lists, with real words, (t84) = -13.77, p = 1.35 x 10^-22, and also with pseudowords, (t84) = -9.42, p = 3.50 x 10^-14, indicating a sentence superiority effect independent of meaningfulness. A direct comparison showed that the advantage of sentence structure was significantly higher within meaningful stimuli (mean d' difference = 1.98, SD = 1.12) than within meaningless stimuli (mean d' difference = 1.36, SD = 0.99; t(52) = 8.58, p = 1.96 x 10^-15). No other significant effects were observed (all F < 0.48, all p > .49), but there was a trending interaction effect between sentence structure and age group, F(1, 51) = 3.84, p = .056, η^2 = .01. Simple effects analysis showed that while performance for sentences was significantly higher in younger adults (M = 3.64, SD = 1.31) than older adults (M = 2.75, SD = 1.40); t(75) = 3.22, p = 7.50 x 10^-3, performance for lists did not differ between younger adults (M = 1.73, SD = 0.73) and older adults (M = 1.33, SD = 0.81); t(75) = 1.47, p = .19. Accordingly, the performance advantage for sentences over lists tended to be larger in younger adults (M = 1.77, SD = 0.79) than in older adults (mean d' difference = 1.29, SD = 0.90); t(50) = 2.07, p = .04. Yet, because the interaction between age group and sentence structure was only marginally significant, this is not strong evidence for an enlarged sentence superiority effect for younger adults.

The trending interaction effect between age group and sentence structure may have also come about because of age differences in vWM capacity (Table 1). Therefore, the composite score of the vWM measures was included as between-subjects covariate into an ANCOVA. The results showed the same main effects for sentence structure, F(1, 50) = 179.39, p = 2.20 x 10^-16, η^2 = .41, and meaning, F(1, 50) = 56.67, p = 8.98 x 10^-10, η^2 = .54, whereas the main effect for age group was not significant anymore. While the interaction effect between sentence structure and meaning, F(1, 50) = 18.12, p = 9.01 x 10^-5, η^2 = .02, persisted, the trending interaction between age group and structure did not continue to differ significantly from zero. No other effects turned out significant (all F < 2.31, all p > .13), indicating no robust disadvantage for older compared with younger adults despite older adults’ vWM limitations.

Finally, the additional post hoc analysis controlling for predictability effects revealed that neither syntactic (median rho across participants = -0.04, quartile 1 [Q1] = -0.27, quartile 3 [Q3] = 0.10, all FDR-corrected p > .99) nor lexical conditional probabilities (median rho across participants = 0.16, Q1 = -0.26, Q3 = 0.30, all FDR-corrected p > .78) could explain the role of the availability of syntactic and semantic constraints for chunking, respectively.

### Discussion

In line with previous studies, Experiment 1 demonstrated an overall lower performance in older adults’ sentences processing. We did not observe a specific benefit of available semantic information during sentence processing in the older adults. Moreover, it remains unclear whether the reduced performance could be associated with specific age differences in syntactic processing because only a trending interaction with age group was apparent. We hypothesized that the nonsignificance of this effect reflects relatively low processing difficulty: possibly, the reliance on syntactic constraints as an encoding strategy to circumvent capacity limits may only become evident when sentence processing is even more capacity-demanding. Therefore, in Experiment 2, the length of the original eight-item stimuli was increased by three items.

### Experiment 2

The overarching aim of Experiment 2 was to further increase the sensitivity in dissociating the respective contributions of syntactic and semantic constraints to age differences in vWM-intensive sentence processing. To this end, vWM demands were increased...
by increasing the sequence length from 8 to 11 items. In addition, interindividual vWM capacity differences were assessed to further dissociate language-specific from domain-general constraints of sentence processing.

**Method**

**Participants.** Participant recruitment followed that of Experiment 1 in all aspects; that is, health, handedness, education, consent, and approval of conduction. From a total of 56 participants, after outlier removal (see Experiment 2 – Statistical analysis), the sample consisted of 26 younger (13 men, age range = 20 – 32 years, mean age = 25.12 years, SD age = 3.24 years) and 27 older adults (13 men, age range = 60 – 70 years, mean age = 64.19 years, SD age = 2.76 years).

**Verbal working memory measures.** As for Experiment 1, all participants underwent the same STM and vWM measures in the same order (Table 1). The scores were then combined into a composite score (α = .73; see Experiment 1 – Statistical analysis).

**Stimuli.** The experimental design of Experiment 2 resembles the design of Experiment 1 (Table 4). However, to increase participants’ reliance on sentence structure and meaning to reduce the design of Experiment 1 (Table 4). However, to increase participants’ reliance on sentence structure and meaning to reduce the design of Experiment 1 (Table 4). However, to increase participants’ reliance on sentence structure and meaning to reduce the design of Experiment 1 (Table 4). However, to increase participants’ reliance on sentence structure and meaning to reduce the design of Experiment 1 (Table 4). However, to increase participants’ reliance on sentence structure and meaning to reduce the design of Experiment 1 (Table 4). However, to increase participants’ reliance on sentence structure and meaning to reduce the design of Experiment 1 (Table 4). However, to increase participants’ reliance on sentence structure and meaning to reduce the design of Experiment 1 (Table 4). However, to increase participants’ reliance on sentence structure and meaning to reduce the design of Experiment 1 (Table 4). However, to increase participants’ reliance on sentence structure and meaning to reduce the design of Experiment 1 (Table 4). However, to increase participants’ reliance on sentence structure and meaning to reduce the design of Experiment 1 (Table 4). However, to increase participants’ reliance on sentence structure and meaning to reduce the design of Experiment 1 (Table 4). However, to increase participants’ reliance on sentence structure and meaning to reduce the design of Experiment 1 (Table 4). However, to increase participants’ reliance on sentence structure and meaning to reduce the design of Experiment 1 (Table 4). However, to increase participants’ reliance on sentence structure and meaning to reduce the design of Experiment 1 (Table 4). However, to increase participants’ reliance on sentence structure and meaning to reduce the design of Experiment 1 (Table 4). However, to increase participants’ reliance on sentence structure and meaning to reduce the design of Experiment 1 (Table 4). However, to increase participants’ reliance on sentence structure and meaning to reduce the design of Experiment 1 (Table 4). However, to increase participants’ reliance on sentence structure and meaning to reduce the design of Experiment 1 (Table 4). However, to increase participants’ reliance on sentence structure and meaning to reduce the design of Experiment 1 (Table 4). However, to increase participants’ reliance on sentence structure and meaning to reduce the design of Experiment 1 (Table 4). However, to increase participants’ reliance on sentence structure and meaning to reduce the design of Experiment 1 (Table 4). However, to increase participants’ reliance on sentence structure and meaning to reduce the design of Experiment 1 (Table 4). However, to increase participants’ reliance on sentence structure and meaning to reduce the design of Experiment 1 (Table 4). However, to increase participants’ reliance on sentence structure and meaning to reduce the design of Experiment 1 (Table 4).

The procedure of Experiment 2 was identical to that of Experiment 1.

**Statistical analysis.** The behavioral measures as well as the statistical analyses in Experiment 2 were identical to those of Experiment 1. In Experiment 2, two younger and one older participant were removed as group outliers (i.e., d’ > 2 SD; there were no below chance performers). Furthermore, RT outliers (for detection criteria, see Experiment 1 – Statistical analysis) and missing responses constituted on average 1.60% (SD = 1.68%) per participant.

**Results and Discussion**

Similar to the results of Experiment 1, the three-way ANOVA (sentence structure × meaning × age group) showed three main effects: sentence structure, F(1, 51) = 181.42, p = 2.20 × 10−16, η2 = .41, meaning, F(1, 51) = 62.66, p = 1.97 × 10−10, η2 = .15, and age group, F(1, 51) = 22.69, p = 1.62 × 10−5, η2 = .19. Figure 2 shows an overview of the single subject and group mean (young vs. old) d’ values for each condition (real word sentences, pseudoword sentences, real word lists, and pseudoword lists).

More precisely, performance for sentences (M = 2.02, SD = 1.12) was higher than for lists (M = 0.85, SD = 0.57). Performance for meaningful stimuli (M = 1.73, SD = 1.19) was higher than for meaningless stimuli (M = 1.14, SD = 0.83). Younger adults performed higher (M = 1.78, SD = 1.13) than older adults (M = 1.09, SD = 0.88). Also, an interaction effect was found between sentence structure and meaning, F(1, 51) = 45.29, p = 1.46 × 10−8, η2 = .11. A simple effects analysis (Bonferroni-corrected) showed that while performance was significantly higher for meaningful sentences (M = 2.55, SD = 1.03) than for meaningless sentences (M = 1.48, SD = 0.94); t(102) = −10.36, p = 5.11 × 10−17, this performance difference was not present for meaningful lists (M = 0.90, SD = 0.62) compared with meaningless lists (M = 0.80, SD = 0.51); t(102) = −0.95, p = 1.00. Moreover, it showed a performance advantage for sentences over word lists, with real words, t(99) = −14.65, p = 6.43 × 10−28 and also with pseudowords, t(99) = −6.03, p = 1.14 × 10−7, indicating a sentence superiority effect independent of meaning. A direct comparison corroborated the suggested larger sentence superiority effect for meaningful (mean d’ difference = 1.66, SD = 0.85) compared with meaningless stimuli (mean d’ difference = 0.68, SD = 0.87); t(52) = 6.78, p = 1.11 × 10−8. In contrast to

Table 4

<table>
<thead>
<tr>
<th>Meaning</th>
<th>Structure</th>
<th>Lists</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real words</td>
<td>der Opa verdarb die Suppe mit dem Salz trotz des Rezepts the granddad ruined the soup with the salt despite the recipe</td>
<td>der verdarb des mit Suppe dem trotz Salz die Opa Rezepts the ruined the with soup the despite salt the granddad recipe</td>
</tr>
<tr>
<td>Pseudowords</td>
<td>der Apo verworbe die Junne mit dem Sohr trotz des Rezebers the Apo ruined the with baby the despite recipe</td>
<td>der verworbe des mit Junne dem trotz Sohr die Apo Rezbers the ruined the with baby the despite baby the Apo recipe</td>
</tr>
</tbody>
</table>

49 AGE DIFFERENCES USING LINGUISTIC CONSTRAINTS
were significant (all $F > 2.58, all p < .11$). Finally, the additional, post hoc analysis assessing predictability effects revealed that neither syntactic (median rho across participants = 0.09, Q1 = −0.13, Q3 = 0.25, all FDR-corrected $p > .76$) nor lexical probabilities (median rho across participants = 0.10, Q1 = −0.14, Q3 = 0.29, all FDR-corrected $p > .77$) could explain the role of the availability of syntactic and semantic constraints for chunking, respectively.

**General Discussion**

Our study reveals age differences in the use of syntactic and semantic constraints for chunking during sentence processing to reduce vWM demands. The results indicate that the availability of syntactic constraints, which enables syntactic chunking (Bonhage et al., 2017; Zaccarelli, Meyer, Makuuchi, & Frederici, 2017), is somewhat less beneficial for older than younger adults (e.g., Feier & Gerstman, 1980; Kemper, 1987; Obler et al., 1991; Stine-Morrow, Ryan, & Leonard, 2000)—particularly when sentences are longer (Experiment 2) but not when sentences are shorter (Experiment 1). Moreover, this age difference persisted even when vWM capacity was accounted for: Older adults displayed a disadvantage in the processing of longer sentences. This indicates difficulties in domain-specific beyond domain-general cognitive abilities. In comparison, we did not find any evidence for an increased recruitment of semantic information in older compared with younger adults. Hence, the benefit of the availability of semantic constraints was comparable across age. Moreover, both age groups showed an advantage of processing meaningful over meaningless sentences, but not lists—increasing the sentence superiority effect by meaningful relations between real words for both age groups.

In general, it is clear that allocating resources to the decoding of syntactic regularities in order to chunk sentences reduces vWM costs—also in the aging population (as indicated by Wingfield, Tun, & Rosen, 1995). Our results, however, suggest that the chunking benefit is reduced in older adults, reflected by a lower benefit of the availability of syntactic regularities when vWM demands are high (Experiment 2). More precisely, only sentences where syntactic regularities were present, but not word lists where syntactic regularities were absent, were remembered less by older than younger adults, suggesting age differences in the sentence superiority effect, in particular in the use of syntactic constraints (as supported by another line of research Craik & Masani, 1967; Gilchrist et al., 2008). This effect was present in spite of a potential emphasis of our serial order task on the memory for word order—and thereby syntactic rather than semantic constraints. While such emphasis could have biased the results toward better syntactic memory, our results suggest that this was apparently not the case. Age differences in syntactic processing persist even when a task focuses on the processing of syntactic constraints (e.g., Gao, Levinthal, & Stine-Morrow, 2012; Stine-Morrow et al., 2010). Therefore, regardless of task-related processes, our results suggest that older adults do not benefit as much as younger adults from syntactic constraints to attenuate their vWM limitations during sentence processing.

Contributing to an ongoing debate, our results show age differences in syntactic processing (e.g., Just & Carpenter, 1992; Kemper, 1986, 1987; Obler et al., 1991; Radvansky et al., 2001;
sentences in a "good-enough" manner (Amichetti et al., 2016; Vansky et al., 2001). As a consequence, older adults may process surface form representations fade away more quickly (e.g., Radev et al., 2000). Increase, older adults allocate fewer resources to the decoding of the life span (e.g., Laver, 2009). Yet, further research is needed to better understand these differences in results.

There are multiple potential explanations for a disadvantage in the use of syntactic constraints for chunking at old age. The most plausible explanation, for us, is that when processing demands increase, older adults allocate fewer resources to the decoding of syntactic regularities (Stine-Morrow et al., 1996); that is, their surface form representations fade away more quickly (e.g., Radyansky et al., 2001). As a consequence, older adults may process sentences in a "good-enough" manner (Amichetti et al., 2016; Christianson et al., 2006; Ferreira, 2003; Traxler, 2014), focusing on the situation, thereby extracting gist rather than detailed knowledge (Tun, Wingfield, Rosen, & Blanchard, 1998). Age differences in the reliance on this processing heuristic appears to persist even when the task context puts particular emphasis on enriched sentence processing (Stine-Morrow et al., 2010), as was the case for the serial order judgment in the present study. Good-enough sentence processing has been proposed to lead to incomplete and shallow representations of syntactic and semantic information (Ferreira & Patson, 2007). The depth to which these representations are processed is associated with the degree of cognitive analysis: higher levels of processing create stronger memory traces (Craik & Lockhart, 1972; Craik & Tulving, 1975). Hence, it is possible that the disadvantage in using syntactic constraints for chunking may arise because good-enough sentence processing leads to shallow representations of syntactic regularities that constitute the basis for the chunking mechanism. If fewer binding resources are allocated to syntactic regularities, thereby creating good-enough rather than detailed syntactic representations, the binding process may become deficient. To this end, binding may be understood as the core basis of chunking: Single elements can be bound into chunks through their strong associations with each other. In support of this hypothesis, it is well known that older adults are generally deficient in binding information to larger entities (Chalfonte & Johnson, 1996; Pandakova, Sander, Werkle-Bergner, & Shing, 2014; Mitchell, Johnson, Raye, & D’Esposito, 2000; Sander, Lindenberger, & Werkle-Bergner, 2012; Sander, Werkle-Bergner, & Lindenberger, 2011; Shing et al., 2010). Our results showed a smaller processing advantage of sentences over lists in older than younger adults, possibly indicating that good-enough processing of syntactic constraints may be insufficient for proper syntactic chunking.

As an alternative to the good-enough processing hypothesis, it is also possible that the diminished advantage of processing sentences over lists, particularly in the more resource-demanding Experiment 2, resulted from an overload of older adults’ vWM capacity associated with increased supraspan interference (Craik & Masani, 1967; Melton, 1963; Wingfield et al., 1995). During sentence processing, supraspan interference can occur when capacity limits are exceeded and syntactic features start to interfere with each other. The ability to suppress interference is a critical determinant of successful sentence comprehension (Glaris, Martin, Van Dyke, Hamilton, & Tan, 2013; Lewis et al., 2006; Santi, Friederici, Makuuchi, & Grodzinsky, 2015; Van Dyke, 2007). Gilchrist et al. (2008) observed that recall performance declined in older adults with increasing number of chunks, particularly so in longer compared with shorter sentences. Similarly, in our study, older adults’ performance was particularly compromised relative to younger adults when sentences were longer (Experiment 2). Therefore, reduced vWM capacity in older compared with younger adults may be associated with increased supraspan interference during sentence processing (Craik & Masani, 1967; Hasher & Zacks, 1988). However, our results do not support this theory given that vWM capacity differences alone cannot explain the age differences in syntactic chunking: When age-related differences in vWM capacity were accounted for, older adults’ disadvantage in syntactic chunking remained considerable. Nonetheless, supraspan interference may not only be subject to vWM capacity limitations but also to age-related differences in the general ability to inhibit interfering material (Hasher & Zacks, 1988). Therefore, future studies should capture the influence of supraspan interference by using additional tasks like the Stroop (Stroop, 1935) or flanker task (Eriksen & Eriksen, 1974).

A third possibility for age differences in the ability to use syntactic constraints for chunking is that older adults may have more difficulties than younger adults in using (syntactic) templates (Charness, 1981; for an overview of template theory, see Gobet & Simon, 1996). Syntactic relationships may be represented as templates, providing structures with fixed slots that have to be filled with variable information (Van Valin & LaPolla, 1997). It is presumed that with experience single elements do not have to be repeatedly chunked into larger entities but at one point can be stored as larger chunks (Gobet & Simon, 1996), like a reoccurring sequence of word categories building a common phrase structure. These common phrase structures may then function as templates. Templates, in general, are thought to be stored in long-term memory (LTM, Gobet & Simon, 1996; Guida, Gobet, Tardieu, & Nicolas, 2012). With syntactic processing being rather automatic, syntactic templates may constitute an implicit, procedural-like long-term memory. Both implicit and procedural memories are relatively spared from age-related decline (e.g., Mitchell, Brown, & Murphy, 1990). Moreover, in our study, we used the same syntactic structure for all sentences. Therefore, a disadvantage in the use of syntactic regularities for chunking is most likely not associated with difficulties in using (the same) syntactic templates (over and over again). However, further research is needed to examine the various options and tackle the specificities of the disadvantage in syntactic chunking.

While the results showed that older adults do not benefit from syntactic constraints as much as younger adults, older adults showed a similar advantage in the use of semantic information when available. That is, younger and older adults both remembered meaningful words better than meaningless pseudowords. This may seem contrary to prior findings showing that older adults have greater difficulties than younger adults in recoding meaningful compared with meaningless words (e.g., Heron & Craik, 1964).
However, our findings additionally show that the advantage in the use of available semantic information was increased in the presence of syntactic constraints. That is, both age groups took advantage of the semantic information of words, in particular when the words were also syntactically related. Therefore, our findings may relate to an enrichment of lexical-semantic representations by relational semantics. This result substantiates theories indicating a role of meaningfulness to syntactic chunking (e.g., Tulving & Patkau, 1962) while refuting the position that there is no interaction between sentence structure and meaningfulness (e.g., Epstein, 1962). In light of age differences in syntactic chunking, the intact ability of semantic processing may become relatively more salient for sentence processing at old age. Therefore, older adults may change their sentence processing strategy toward a semantic approach (Amichetti et al., 2016). However, further research is needed to substantiate this suggestion.

**Conclusion**

Our study suggests that sentence processing is facilitated by both the ability to use syntactic constraints for chunking and the availability of semantic information. While both younger and older adults profit from semantic information to a similar extent, the benefit of using syntactic constraints for chunking decreases with normal aging. Hence, the relative importance of semantic information processing is increased for successful language comprehension in old age. This finding persisted even when individual differences in vWM capacity were accounted for, pointing to a disadvantage in language-specific beyond domain-general cognitive abilities in old age. The salience of semantic information may lead to a change in sentence processing strategies toward a semantic approach later in life. Further research endeavors are needed to support this hypothesis.

**References**


