J Psycholinguist Res
DOI 10.1007/s10936-011-9197-8

Working Memory Effects of Gap-Predictions in Normal Adults: An Event-Related Potentials Study

Arild Hestvik · Evan Bradley · Catherine Bradley

© Springer Science+Business Media, LLC 2012

Abstract The current study examined the relationship between verbal memory span and the latency with which a filler-gap dependency is constructed. A previous behavioral study found that low span listeners did not exhibit antecedent reactivation at gap sites in relative clauses, in comparison to high verbal memory span subjects (Roberts et al. in J Psycholinguist Res 36(2):175–188, 2007), which suggests that low span subjects are delayed at gap filling. This possibility was examined in the current study. Using an event-related potentials paradigm, it was found that low span subjects have an onset latency delay of about 200 ms in brain responses to violations of syntactic expectancies after the gap site, thus providing a time course measure of the delay hypothesized by previous literature.

Keywords Gap-filling · Working memory · Sentence comprehension

Introduction

In a recent paper, Roberts et al. (2007) examined whether working memory differences among adults and children resulted in different patterns of antecedent reactivation in filler-gap constructions. They used the cross-modal picture priming method (Love and Swinney 1997; Love 2007) to test for immediate antecedent reactivation of relativized noun phrases in double object constructions, where the gap was non-adjacent to the verb. To illustrate, they presented auditory sentences like (1)

(1) John saw [the peacock]$_i$ to which the small penguin gave the nice birthday present t_i in the garden last weekend.

A. Hestvik (✉) · E. Bradley
Department of Linguistics and Cognitive Science, University of Delaware, Newark, DE, USA
e-mail: arild@udel.edu

C. Bradley
Project on Language and Spatial Development, Florida International University, Miami, FL, USA

Published online: 17 January 2012

They found that subjects with high verbal memory span exhibited antecedent reactivation priming at the gap-position t_i in a cross-modal picture categorization task, but low verbal memory span subjects did not show evidence of antecedent reactivation at the same syntactic position. The authors suggest that the low-span population require "more time to integrate the filler at the gap position and reactive the filler only later during the sentence." This predicts that if antecedent reactivation was measured at a later point in time, it would then be observed also for the low verbal memory span subjects.

A question left open by Robert et al.'s study is: how delayed are low verbal memory span subjects at gap-filling, and how measure late gap-filling? Without a specific temporal prediction, use of the cross-modal picture priming method would require sampling reactivation at multiple down-stream syntactic positions, which would be a resource intensive endeavor given the large subject samples, small effect size, repeated measures and multiple sessions required for a single position. Furthermore, although the Roberts et al.'s findings is consistent with delayed gap-filling in low span subjects, their observations would also be consistent with a complete lack of syntactic gap-filling—a possibility that is theoretically unlikely (and would require low span subjects to interpret filler-gap dependencies via alternative mechanisms). In any case, what is needed to conclude that low span subjects are delayed at gap-filling is positive evidence about *when* they actually do fill the gap.

Another possible problem with interpreting the results of Roberts et al. (2007) is that low verbal memory span in the subjects could in fact be a confound, given the experimental methodology. Cross-modal picture priming is a dual-attention task, and the premise of the method is that trials are only valid in so far as the subjects are paying equal attention to stimuli in both modalities. Paying equal attention to two stimulus streams raises the baseline amount of working memory that has to be allocated during the task. This raises the possibility that if a subject has low verbal memory capacity, he or she may in fact not be able to pay equal attention to both stimulus stream in every trial, focusing more on either the auditory stimuli or the picture stimuli. Indeed, the Listening Span test of Daneman and Carpenter (1980), which was used by Roberts et al., measures precisely the ability to keep two types of information in mind at the same time. If so, it could be that low verbal memory span subjects fail to show priming because they are unable to comply with dual task demands. It is therefore desirable to explore additional methods for measuring the effect of working memory resources on processing of filler-gap constructions, where the task itself does not add extra working memory demands.

To summarize, two issues are left open by the Roberts et al. study: first, if low verbal memory span subjects are delayed at gap-filling, how much are they in fact delayed? Second, how can gap-filling be measured in low-span subjects without letting their low verbal memory span potentially get in the way of the measurement itself?

The Current Study

We addressed both issues by using the "filled gaps" event-related brain potentials measure of gap-filling (Hestvik et al. 2007). This paradigm examines whether listeners fill a gap by measuring the latency of automatic brain responses to expectancy violations. Specifically, it measures event-related potentials to violations of on-line expectations about syntactic categories and structure immediately *following* a predicted gap position. In this paradigm, listeners are presented with ungrammatical expressions like "The zebra that the hippo had kissed *the camel* on the nose…" If listeners predict and posit a gap immediately after *kissed*, the probability of encountering a noun phrase in the same position is low or zero. In our previous study using this paradigm (Hestvik et al. 2007), the category expectancy violation elicited an early Left Anterior Negativity (eLAN) peaking between 100–200 ms after onset of *the*. The fast response to *the* shows that the parsing system already predicted and filled a gap in this position, and accounts for why the expectancy response is so early. Furthermore, the measure serves as a temporal index of the gap-filling process. If a group of subjects are delayed in their gap-filling, this would then be reflected in a temporally delayed brain response; similarly, if a group of subjects fail to fill gaps, it would be reflected in the absence of a brain response to the expectancy violation (because no expectancies were generated).

This paradigm is well suited for measuring whether gap-filling latencies differ between two populations of subjects. Furthermore, it is well suited to low verbal memory span subjects, because it only involves passive listening to auditory stimuli, followed by comprehension questions about the sentences. The task requires no more than normal language processing abilities, and involves no secondary attention task. Hence, there will be no task-related confounds related to limitations in verbal memory span.

Method

Subjects

52 college-aged subjects volunteered and participated in the study. Following the same methodology as Roberts et al., participants were required to completed the Listening Span Test of Daneman and Carpenter (1980). (5 subjects were excluded due to not completing the Listening Span test, and one subject was excluded due to only completing half the experiment, leaving 46 subjects.) Verbal memory span was computed as in Daneman and Carpenter (1980), with the addition that if a subject got one correct answer at any level higher than their scores computed as in Daneman and Carpenter (1980), we added 0.5 to their score. This had the effect of providing a better overall estimate of span, and served to stretch out the scale and provide a better differentiation between groups. By this method, the median verbal memory span of all the subjects was 3, as in Roberts et al.'s study. Following that study, we assigned subjects to the low span group if their span was less than or equal to the median, and to the high span if it was greater than the median. This resulted in 19 high span subjects (mean age 21, $SD = 2$; 14 females) and 27 low span subjects (mean age $= 21$, $SD = 4$; 18 females). The imbalance of gender reflects the student body that we sampled from.

The mean WM score in the high span group was 3.6 ($SD = .52$), and the mean WM score in the low span group was 2.4 ($SD = .47$). There was no significant difference in WM scores between the male and female subjects ($F(1, 44) = .24$, $p = .6$). Post-experiment analysis revealed a weak but significant positive correlation between WM score and accuracy on the comprehension questions following the stimuli ($r = 0.46$, $p < .05$), and a one-way ANOVA of the accuracy dependent variable revealed a significant difference between the two WM groups ($F(1, 44) = 14.6$, $p < .001$. This post-hoc analysis further bolsters the validity of the WM span test and the bifurcation of subjects into low and high span groups.

Stimuli and Design

The stimuli consisted of four sentence types, illustrated in Table 1, comprising four experimental conditions.

The TRACE condition was used to prevent participants from predicting that each relative clause would turn out to be ungrammatical, and the OBJECT condition prevented participants from always expecting a relative clause. The critical comparison in the experiment

Table 1 Examples of sentence in each experimental condition

Condition	Sentence
UNGRAMM	The zebra that the hippo kissed **the camel* on the nose ran far away
ADJUNCT	The weekend that the hippo kissed *the camel* on the nose he ran far away
OBJECT	The zebra said that the hippo kissed the camel on the nose and then ran far away
TRACE	The zebra that the hippo kissed on the nose ran far away

The critical items for comparison between the UNGRAMM and ADJUNCT conditions are italicized

was between *the camel* in the ungrammatical UNGRAMM condition, and the same noun phrase in the ADJUNCT condition. Sentences in these two conditions have the same syntactic structure and identical strings leading up to the critical noun phrase; the difference is that the critical noun phrase is consistent with expectations in the ADJUNCT condition, but unexpected in the UNGRAMM condition. This is due to the difference in grammatical function of the two relativized nouns. The gap in ADJUNCT sentences is predicted to be found *after* the post-verbal noun phrase, whereas in the UNGRAMM condition, the gap is predicted to occur *before* the same noun phrase—thus making the presence of that noun phrase highly unexpected.

The stimuli were constructed by first making a list 32 sentences in the form of the grammatical TRACE condition (well-formed relative clauses). For each sentence in the TRACE list, a corresponding sentence in the UNGRAMM, ADJUNCT and OBJECT condition was constructed, resulting in a total of four lists of 32 sentences each. A second set of four lists of 32 sentences were then constructed by switching the agent and patient noun phrase in each original sentence quadruplet with that of another sentence, as well as making other changes to the post-verbal continuation part of the sentence. This resulted in 64 sentences in each condition, for a total of 256 stimuli sentences.

The two sets of four lists were labeled Script A and Script B respectively. Half the subjects were presented with the four lists in Script A in the first block, and the second set of four lists in Script B in the second block. The other half were presented with Script B in the first block and Script A in the last block. The order of presentation within each list was randomized on-line for each subject, and the order of selection from each condition list was also randomized on-line.

Each sentence was matched to a single unique comprehension question, asking about the stimulus sentence just heard. The 64 comprehension questions where equally distributed among four types: object Wh-questions (“Who did the alligator tap?”), subject Wh-questions (“Who bumped the duck?”), Yes-No questions (“Did the hippo kiss the camel?”) and a set of “easy” non-content Yes-No questions (“Did you hear the word ‘road’?”). Question type was counterbalanced with experimental condition type of the stimulus sentences. Each question was matched with two picture response options. One picture represented an object or character; the other picture represented a question mark. Subjects were instructed to select the depicted object if it represented the correct answer, or the question mark if it did not. Half the trials presented a picture depicting the correct answer, the other half required the choice of the missing answer. All the UNGRAMM sentences were matched to the missing answer option (so as not to ask a comprehension question about an ungrammatical sentence). An example of a stimulus sentence and the corresponding question is given in Fig. 1.

The stimuli were digitally recorded using 16 bit resolution and 22,050 kHz sampling rate. The sentences were read by a professional linguist, and intonation was controlled by

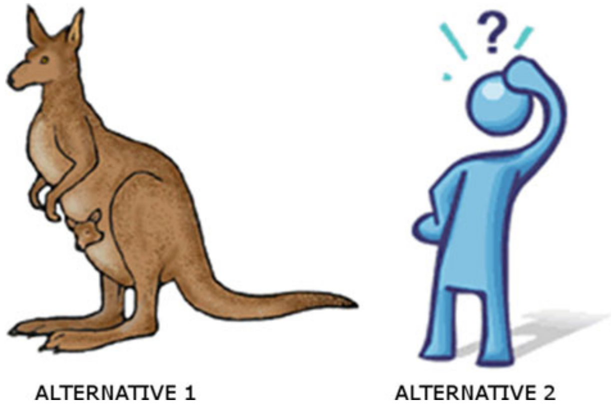

Fig. 1 Example of a stimulus sentence and comprehension question, with answer choices and correct answer. Stimulus: "The winter that the ostrich raced the giraffe down the road, it was humid!" Question: "Who did the ostrich race?" (Alternative 2 is the correct answer)

first reading the OBJECT sentence with a neutral intonation, and then imposing the identical prosodic contour on the substring starting with the critical underlined noun phrase in the UNGRAMM and ADJUNCT condition. In this way, the auditory stimuli mimicked orthographically represented stimuli in being stripped of prosodic cues to gap-filling (Nagel et al. 1994). The comprehension questions were read by a different female speaker. The average number of words per sentence was 14.8 (SD = 2), and the average duration of each word was 326 ms (computed by dividing sentence length into number of words), yielding a speaking rate of about 3 words/s, or about 180 words/min (corresponding to conversational speech rate).

Differing from Hestvik et al. (2007), we excluded additional filler sentences in order to shorten EEG recording session durations. Thus, the proportion of ungrammatical sentences was 25%, which is higher than in the previous study (∼17%). Based on the finding by Hahne and Friederici (1999) that early ERPs reflecting first-pass parsing operations (eLAN and LAN) were elicited by both high (80%) and low (20%) proportion of ungrammatical sentences in that study, we did not expect this higher proportion of ungrammatical sentences to affect the early ERPs, although it could affect the P600, as found by Hahne and Friederici.

Procedure

Subjects listened to sentences presented in free field speakers. After every sentence, a comprehension question was presented, which required an answer selected by pressing one of two buttons (corresponding to one of two picture choices presented on an LCD screen; one picture was either correct or incorrect, the other picture signified that the subject decided that no correct answer was presented on the screen). EEG was recorded with a 128-channel EGI 300 system in a sound shielded audiometric booth, with sampling rate of 250 Hz, and an analog lowpass filter at 100 Hz. The experiment was controlled by E-Prime software (Schneider et al. 2002). The 256 stimuli sentences were presented in pseudo-randomized order. The entire recording session took about 1 h 15 min.

To measure expectancy about syntactic category and structure after the putatively posited post-verbal gap, ERPs were time-locked to the onset of the article of the ungrammatical "filled gap" noun phrase in the UNGRAMM condition and the grammatical noun phrase in the ADJUNCT condition. If participants predicted a gap after the verb in the UNGRAMM condition, then encountering the unexpected noun phrase should elicit a syntactic category expectancy violation ERP. Based on our own previous results and the empirically based model of Friederici (2002), we predicted an Early Left Anterior Negativity to the structural category violation. Furthermore, the P600 is an index of syntactic integration difficulty (Kaan et al. 2000). The presence of a P600 in the current data is expected because the ungrammatically filled gap NP cannot be integrated with the analysis of the verb phrase as the direct object function has already been assigned to the trace.

After recording, single trial epochs were computed starting from the onset of the article *the* introducing the critical noun phrase, including a 400 ms baseline period before this time point. Trials with eye blinks and eye movement artifacts were then removed, and channels where the maximum difference during the epoch was greater than 200 mV was marked as bad and replaced with the spline interpolation. Each epoch was then baseline corrected based on the 400 ms period preceding the critical noun phrase, and each subjects' trials in each condition were then averaged and re-referenced to the average voltage of all electrodes.

Results

Behavioral Results

The mean accuracy on comprehension questions was 80% (SD = 5%, range 68–93%). Note that this includes answers to questions about the UNGRAMM sentences (which always required the "missing correct answer" response); without this condition overall accuracy was 86%. By group, low span subjects' mean accuracy was 78% (SD = 5%), and high span subjects mean accuracy was 83% (SD = 3%).

In order to assess whether the two groups differed significantly in comprehension accuracy, and whether any such difference interacted with the experimental condition of the stimulus sentences, the mean proportion correct answers per subject and condition was submitted to a mixed factorial repeated measures ANOVA with WMGROUP (high vs. low verbal working memory span) as the between subject variable, and CONDITION (4 levels) as the within subjects variable. This revealed a main effect of WMGROUP ($F(1, 44) = 14.17$, $p < .001$) and a main effect of CONDITION ($F(3, 132) = 207.5$, $p < .0001$), cf. Fig. 2, left panel. The interaction term CONDITION × WMGROUP was not significant ($F(3, 132) = 1.49$, $p = .22$). Post-hoc Scheffé test showed that the UNGRAMM condition's accuracy was lower than all other conditions, and that the OBJECT condition had lower accuracy than the ADJUNCT and TRACE condition; the TRACE and ADJUNCT conditions (the two grammatical relative clauses) did not differ. The correlation between each subject's verbal memory span score and overall accuracy was also computed, which revealed a weak but significant correlation between verbal memory span and comprehension question accuracy ($r = .46$, $r^2 = .21$, $p < .005$), cf. Fig. 2, right panel.

ERP Results

After artifact decontamination, four high span subjects and six low span subjects were removed from data analysis because their total number of valid trials was less than 50% of the total. For the remaining 36 subjects (N = 21 for low WM and N = 15 for high WM), a one-way ANOVA confirmed that there was no effect of group on the number of good trials. In fact, the mean number of good trials for the low span group was slightly higher

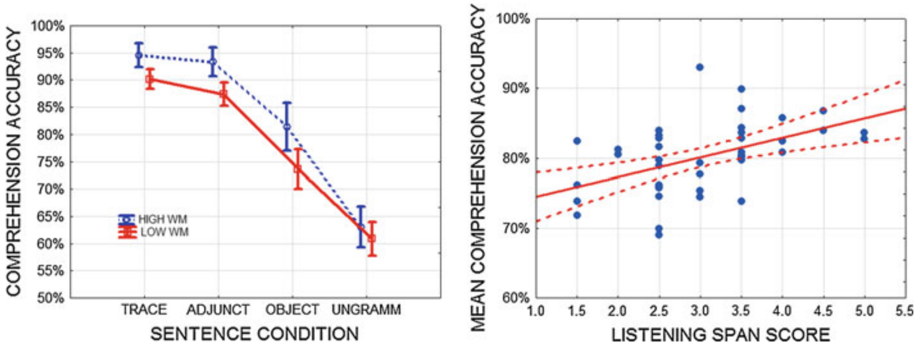

Fig. 2 *Left panel*: comprehension accuracy by condition and WM group (vertical bars denote 95% confidence intervals). *Right panel*: linear relationship between listening span score and mean comprehension accuracy. Dotted line indicates 95% confidence interval

(M = 80%, SD = 11%) than for the high span group (M = 78%, SD = 16%). Thus, statistical power based on number of trials would be the same for both groups. In addition, the same statistical results regarding the behavior data held true for this slightly smaller sample (effect of WMGROUP on accuracy: $F(1, 33) = 7.27$, $p < .05$; effect of CONDITION: $F(3, 99) = 164.0$, $p < .0001$), no interaction CONDITION × WMGROUP).

In order to inspect the main grammaticality effect, difference waveforms of UNGRAMM-ADJUNCT was computed for all subjects and the mean difference was plotted as topographical map over time as in Fig. 3. This shows the temporal and spatial distribution of the difference between the UNGRAMM and ADJUNCT control condition. An early anterior negativity to the UNGRAMM condition compared to the ADJUNCT condition can be observed bilaterally in both left and right anterior inferior electrodes, and a later positivity to the UNGRAMM condition can be observed in the central to right posterior dorsal electrodes.

In order to analyze these effects, we computed the mean voltage of the left and right anterior inferior electrode regions (both were constructed in order to test whether the anterior negativity differed on the left and right side); as well as the mean voltage of a sample of electrodes in the central-to-right posterior dorsal region carrying the late positivity. The electrode regions are highlighted in Fig. 4.

Left Anterior Negativity

The mean voltage of the left and right anterior inferior region was computed for each 100 ms time window between 0 and 1000 ms, for each subject and condition, and submitted to a repeated measures mixed factorial ANOVA, with WMGROUP as between-subject factor and CONDITON (2) × HEMISPHERE (2) × TIME (10) as the within-subject factors. (Here and in the following we only report statistics involving the CONDITION factor). This revealed a main effect of CONDITION ($F(1, 34) = 18$, $p < .001$) such that UNGRAMM was more negative than ADJUNCT; a marginal CONDITION × WMGROUP interaction ($F(1, 34) = 3.14$, $p = .085$), and a marginally significant 3-way interaction TIME × CONDITION × WMGROUP (using Greenhouse-Geisser correction) ($F(9, 306) = 2.4$, $\varepsilon = 0.38$, $p = .062$). As there was no effect of hemisphere in the omnibus ANOVA, the laterality factor was collapsed. We next tested in which specific time windows the CONDITION effect was significant. Because we had predictions about a latency difference between groups in the anterior negativity effect and planned to compare the two conditions over time, orthogonal

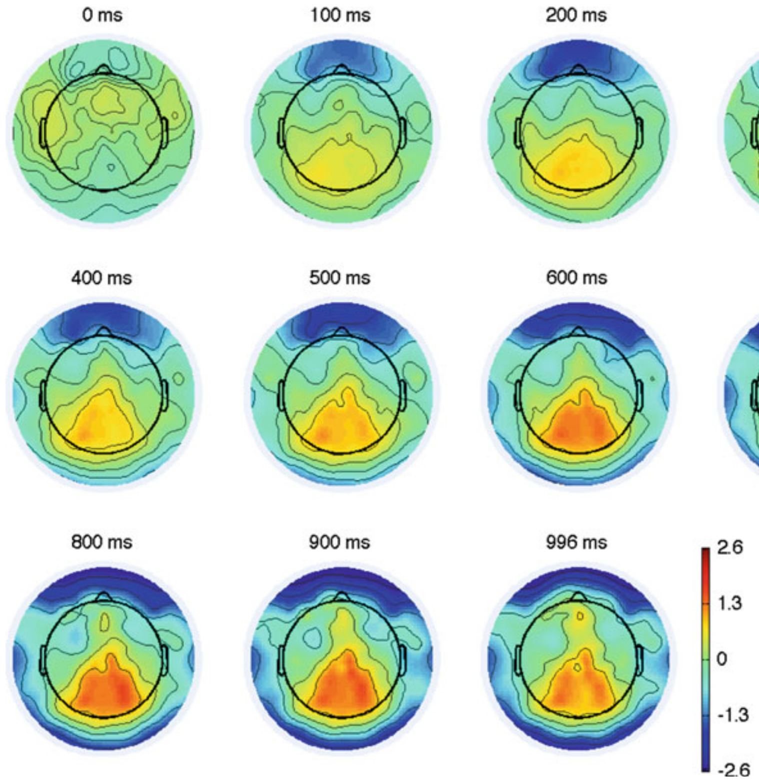

Fig. 3 Topomap time series of the difference wave constructed by subtracting ADJUNCT from UNGRAMM

contrast analyses was employed to compare the UNGRAMM versus ADJUNCT difference in successive time windows for each WM group. (For all orthogonal contrasts reported below, the degrees of freedom was 34, and the reported probabilities are two-tailed).

In the low span group, the difference was significant during the 600–700 ms time window ($t = 2.6$, $p < .05$) and in the subsequent 700–800 ms and 800–900 ms time windows. In the high span group, the difference was significant already during the 400–500 ms time window ($t = 2.42$, $p = .02$) and every subsequent time window until the end of the epoch. Thus, the bilateral anterior negativity started 200 ms earlier for the high span group. This difference in onset latency for the anterior negativity is illustrated in Fig. 5. As shown in Fig. 5, there is also an earlier peak in the anterior negativity data. The peak latency of this waveform is at 180 ms, a time course consistent with previous findings of early Left Anterior Negativity (eLAN) (Hahne and Friederici 1999; Hestvik et al. 2007).

In order to analyze this effect specifically, and given the fact that the eLAN typically has a short duration, we analyzed the 150–200 ms time window around the peak latency for the same bilateral electrode region and tested for significance of the grammaticality effect and whether it interacted with group with a mixed factorial repeated measures ANOVA. This resulted in a significant CONDITION × WMGROUP interaction ($F(1, 34) = 5.0$, $p = .03$) for the (bilateral) eLAN, illustrated in Fig. 6.

To summarize, both verbal memory span groups showed a bilaterally distributed anterior negativity effect for the ungrammatically filled gap NP (which we interpret as functionally

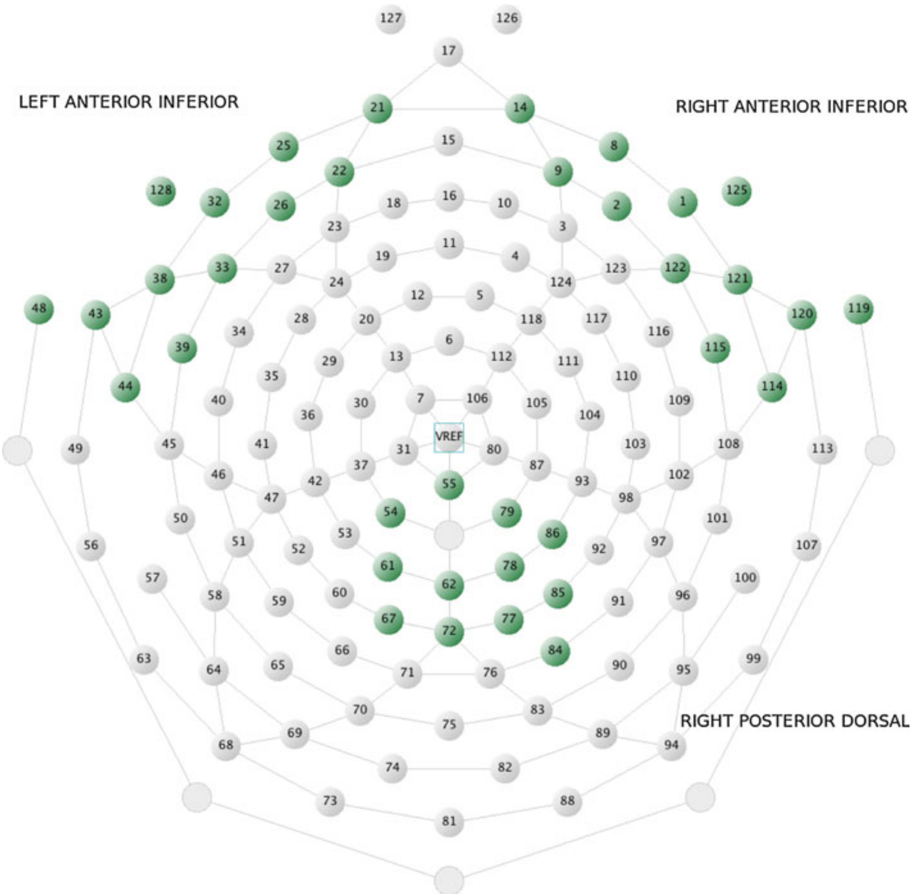

Fig. 4 Electrode regions used for averaging electrodes for the LAN and the P600 effect

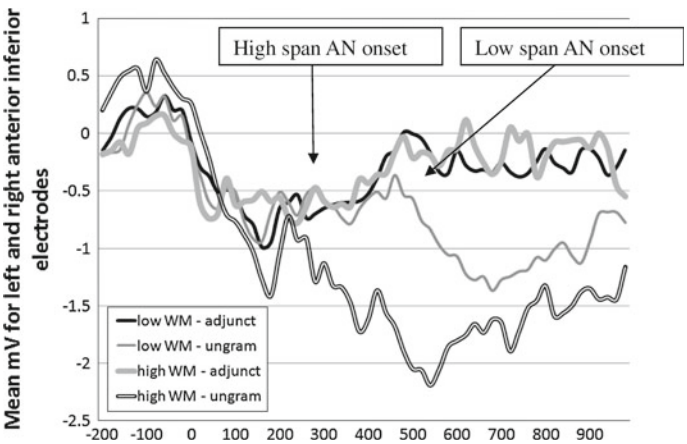

Fig. 5 Anterior negativity effect of filled gap NP, mean of left and right anterior inferior electrode regions. High span subjects exhibit earlier onset latency and peak of LAN than low span subjects (waveform smoothed form display)

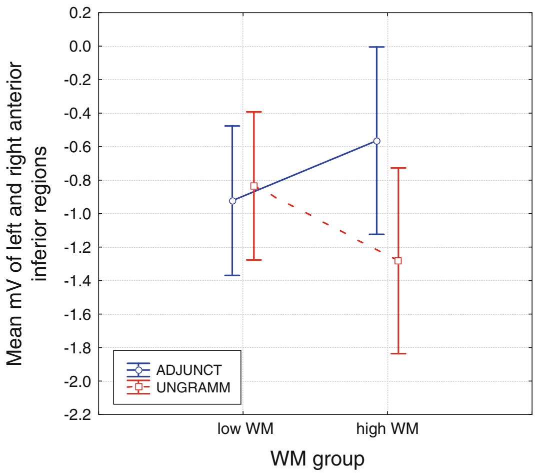

Fig. 6 Interaction plot of CONDITION × WMGROUP

equivalent to a left-anterior negativity), but only the high span group showed an early AN (interpreted as functionally equivalent to the eLAN).

P600

A clear positivity to the UNGRAMM condition was observed in the posterior region between 400–1000 ms, cf. Fig. 1. This effect was interpretable as the P600 ERP component. Visual inspection revealed that the distribution of the P600 was slightly right-lateralized. This distribution was objectively confirmed by a spatial Independent Component Analysis (ICA) (Bell and Sejnowski 1995) of the UNGRAMM minus ADJUNCT difference waveform data. The ICA analysis was conducting using the ERP PCA Toolbox v.2.23 (Dien 2010). 18 spatial factors was retained (based on the parallel scree plot test, cf. (Horn 1965)) and rotated with the Infomax rotation, accounting for 78% of the variance. The first spatial factor, accounting for 10% of the variance was identified as corresponding to the P600. As seen in Fig. 7, the distribution is slightly right-lateralized.

For this reason, we sampled the three posterior midline electrodes, three electrodes on each side of the midline, and an additional three electrodes to the right of the midline (cf Fig. 4), to achieve representative spatial sampling. The mean voltage of this electrode set was computed and used as the dependent measure in a mixed factorial repeated measures ANOVA. 10 consecutive 100 ms time windows for each condition and subject was computed and submitted to a mixed factorial repeated measures ANOVA, with WMGROUP as the between-subjects factor and CONDITION (2 levels: ADJUNCT and UNGRAMM) and TIMEBIN (10 levels) as the within-subject factors. The ANOVA revealed a main effect of CONDITION ($F(1, 34) = 21.68$, $p < .0001$), a significant TIME × CONDITION interaction ($F(9, 306) = 20.13$, $\varepsilon = 0.37$, $p < .0001$) and a 3-way TIME × CONDITION × WMGROUP interaction ($F(9, 306) = 2.68$, $\varepsilon = 0.37$, $p < .05$). The TIME × CONDITION interaction was caused by the fact that the difference between UNGRAMM and ADJUNCT was not the same in every time window, i.e., the effect developed over time (which is expected for an ERP). The 3-way interaction indicates that the P600 effect became significant at different latencies for the two WM groups. The mean waveform in Fig. 8 for the selected electrode region illustrates this interaction.

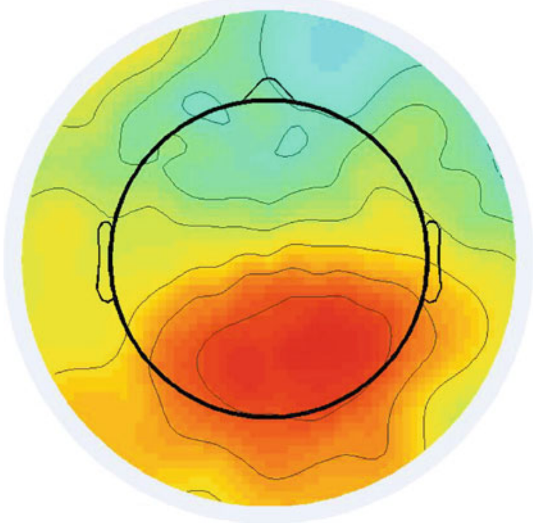

Fig. 7 Spatial distribution of independent component for the P600 effect

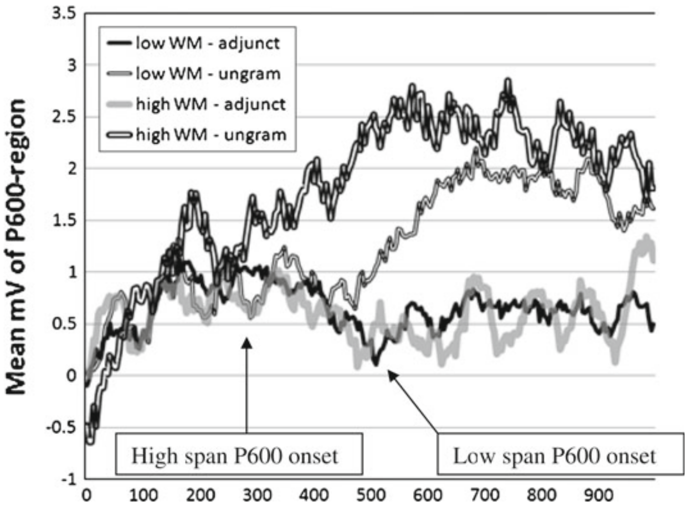

Fig. 8 Effect of WM group on P600. Arrows indicate approximate onset latency of P600 effect (the positive voltage of the ungrammatical condition compared to the control condition); the high WM group's onset latency is about 200 ms earlier than the low WM group

In order to determine the onset of the P600 wave for each group, the 3-way interaction was followed up with orthogonal contrast analysis of successive 100 ms time windows, testing for significance of UNGRAMM versus ADJUNCT difference separately for each WM group. For the low span group, the effect of CONDITION became significant in 600–700 ms time window ($t = -2.8$, $p < .01$) and was significant in every successive time window. For the high span group, the effect was significant in the 400–500 ms time window ($t = -2.25,\ p = .03$), and in every subsequent time window. Thus, the P600 was significant 200 ms earlier for the high span group in comparison to the low span group.

Discussion

Behavioral Results

Verbal memory span predicted the degree of off-line comprehension of the stimuli sentences. This is expected, because in order to correctly answer a comprehension question, the previous stimulus sentence must be kept in and then retrieved from working memory, and truth conditions for the additional question sentence must computed relative to the stimulus sentence. This means that two sentences must be kept in working memory at the same time. This predicts that subjects with low verbal memory span should perform poorer than high span subjects on this task. Indeed, the significant difference in comprehension accuracy between low span and high span subjects and the related correlation between individual memory span and comprehension accuracy confirmed that the Listening Span task provides a valid partitioning of the subjects into two distinct groups. Furthermore, the lack of interaction between experimental condition effect on accuracy and memory span group, shows that low verbal memory span did not impair the low span subjects' ability to comprehend sentences with filler-gap dependencies. I.e., the lack of interaction between condition and span group on accuracy suggests that low-span subjects were able to compute filler-gap dependencies.

ERPs

Both low and high span subjects exhibited a bilateral Anterior Negativity in the 500–800 ms time range, to the filled gap noun phrase. The time course of this AN is similar to that of LAN effects reported in the literature, with AN peaking for the high span group at about 500 ms. The current experiment was based on a similar experiment reported in Hestvik et al. (2007), where a left-lateralized anterior negativity was found. We expected to observe the same left-lateralized negativities, and the finding of bilaterally distributed anterior negativity was therefore not expected. However, we note that the left-lateralized effect in Hestvik et al. (2007) was only observed after the data was decomposed into temporo-spatial PCA components, and the undecomposed anterior negativity was in fact bilateral. Another difference was that no P600 was observed in Hestvik et al. (2007)—unexpected in that study—whereas a clear P600 was observed in the current study. Possibly, the slightly different ratio of ungrammatical sentences and the less varied stimuli (by the omission of filler sentences) in the current study is responsible for these differences. Nevertheless, we feel confident in interpreting the anterior negativity as reflecting the same underlying mechanisms that give rise to left-lateralized negativities, as many other studies reported in the literature has observed bilaterally distributed anterior negativities were left-lateralized effects were expected (Muller et al. 1997; Hagoort et al. 2003; Ueno and Kluender 2003; Phillips et al. 2005; Sabisch et al. 2009).

Turning to the latency differences, the high span subjects' AN became significant in the 400–500 ms time window, whereas it only became significant during the 600–700 ms time window for the low span subjects. This is consistent with the speculation of Roberts et al. (2007) that low span subject "need more time to fill a gap." In addition, the high span group also exhibited a separate peak in the AN at around 180 ms; although bilateral, this is partially consistent with the findings of an early Left Anterior Negativity to the same manipulation in Hestvik et al. (2007) and will be interpreted as functionally the same response. The lack of an early AN in the low span group is also consistent with a delayed gap-filling response, and suggest that the low span group's parsing operations are less automatic.

Table 2 Time course differences of structure building, gap-filling and ungrammaticality detection on the two memory span groups, and resulting ERP index time course differences

Latency (ms)	Low span ERP onset	High span ERP onset
100–200	–	Early AN
200–300	–	AN
400	AN	P600
600	P600	

The P600 response similarly follows the same delayed time course for low span subjects. For the high span group, the P600 onsets at around 300 ms and peaks between 600 and 700 ms. For the low span group, the P600 onsets at around 500 ms and peaks between 700 and 900 ms (cf. Fig. 5). According to Kaan et al. (2000), a P600 is elicited when a syntactic category cannot be integrated with the analysis of the current structure, which indeed is the case for the filled gap noun phrase. The delayed index of integration difficulty in low span listeners results because the integration difficulty only becomes an issue after a gap has been posited with a 200 ms delay.

Conclusion

We summarize the inferences about time course differences of incremental syntactic structure building and gap-filling in the two groups in Table 2. In conclusion, the current findings provide support from electrophysiological measures of gap-filling that low verbal memory span subjects are slightly delayed in their construction of filler-gap dependencies. This hypothesis was raised by the results of a behavioral antecedent reactivation experiment (Roberts et al. 2007), where low span subjects failed to show immediate reactivation of fillers at gap sites. In the current study, it was shown that low span subjects are about 200 ms delayed at exhibiting brain responses reflecting the detection of gap-inconsistent structure. This provides converging evidence that low span subjects have not constructed a gap yet at the temporal juncture sampled by the method in Roberts et al. (2007); and explains why they did not exhibit reactivation priming at this point in time. However, 200 ms later this group must have constructed a gap in order to generate the delayed sequence of AN and P600 responses. This predicts that if Roberts et al. (2007) had tested for antecedent reactivation 200 ms downstream from the gap position in their study, low span subjects should then exhibit antecedent reactivation.

Acknowledgments The current research was supported by a UNIDEL grant from the University of Delaware to the author. We thank our lab assistants Catherine Bradley, Tyler Prescott, Megan Kaufmann, Lauren Sparacino and Amanda Acca for assisting with the data collection for the current study and the University of Delaware undergraduates who volunteered as research participants. Stimulus materials are available from author upon request.

References

Bell, A. J., & Sejnowski, T. J. (1995). An information-maximization approach to blind separation and blind deconvolution. *Neural Computation, 7*(6), 1129–1159.

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

Dien, J. (2010). The ERP PCA toolkit: An open source program for advanced statistical analysis of event-related potential data. *Journal of Neuroscience Methods, 187*(1), 138–145.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences, 6*(2), 78–84.
Hagoort, P., et al. (2003). Syntax-related ERP-effects in Dutch. *Cognitive Brain Research, 16*(1), 38–50.
Hahne, A., & Friederici, A. D. (1999). Electrophysiological evidence for two steps in syntactic analysis. Early automatic and late controlled processes. *Journal of Cognitive Neuroscience, 11*(2), 194–205.
Hestvik, A., et al. (2007). Brain responses to filled gaps. *Brain and Language, 100*(3), 301–316.
Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*, 179–185.
Kaan, E., et al. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes, 15*(2), 159–201.
Love, T. (2007). The processing of non-canonically ordered constituents in long distance dependencies by pre-school children: A real-time investigation. *Journal of Psycholinguistic Research, 36*(3), 191–206.
Love, T., & Swinney, D. (1997). *Real-time processing of object-relative constructions by pre-school children*. Los Angeles, CA: 10th Annual CUNY Conference on Human Sentence Processing.
Muller, H. M., et al. (1997). Event-related potentials elicited by spoken relative clauses. Brain research. *Cognitive Brain Research, 5*(3), 193–203.
Nagel, H. N., et al. (1994). Prosody and the processing of filler-gap sentences. *Journal of Psycholinguistic Research, 23*(6), 473–485.
Phillips, C., et al. (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research, 22*(3), 407.
Roberts, L., et al. (2007). Antecedent priming at trace positions in children's sentence processing. *Journal of Psycholinguistic Research, 36*(2), 175–188.
Sabisch, B., et al. (2009). Children with specific language impairment: The role of prosodic processes in explaining difficulties in processing syntactic information. *Brain Research, 1261*, 37–44.
Schneider, W., et al. (2002). *E-prime reference guide*. Pittsburgh: Psychology Software Tools Inc.
Ueno, M., & Kluender, R. (2003). Event-related brain indices of Japanese scrambling. *Brain and Language, 86*(2), 243–271.