Individual Differences in the Expectation-based Comprehension of Korean Dative Sentences: An ERP Study

Hongoak Yun[1*], Dongsu Lee[2], Yunju Nam[2]

[1]*Jeju National University*

[2]*Konkuk University*

Abstract

Readers' active use of linguistic cues from a given context elicits anticipatory processing of yet-to-be-encountered information. In this study, we aimed to examine whether the patterns of anticipatory comprehension would systematically differ by the degree of readers' working memory capacity. Readers' evoked responses potentials (ERPs) in response to words were recorded in the processing of Korean dative sentences (i.e., subject+[recipient+theme]/[theme-recipient]+adverb+verb) in which the presentation order of arguments (i.e., role predictability) and the likelihood corresponding to argument role fillers (i.e., word predictability) were manipulated. We found quantitative and qualitative differences in ERPs among readers during sentence comprehension. The N400 emerged in the integration of unpredictable words, and it occurred more frequently among readers with low working memory. Of our interest, we observed the asymmetrical distribution of the negativity and the positivity, attributable to the differences in readers' working memory, at adverbs and verbs in which readers were busy with integrating previously-presented arguments into sentences and processing incoming words. Our results suggested that readers with low working memory are more involved in the lexical retrieval process, whereas those with high working memory are more attentive to the structural or semantic integration process. In short, we argued that the lack of working memory capacity could make readers fall

Received 12th December 2021; Revised 3rd April 2022; Accepted 5th April 2022
* Correspondence: hongoakyun@jejunu.ac.kr
Journal of Cognitive Science 23(1): 57-94 March 2022

behind in using lexical and structural information during sentence processing, particularly for argument integration.

Key words: *word predictability, role predictability, Korean dative sentences, individual differences, working memory capacity, expectation*

1. Introduction

It has widely accepted that readers integrate incoming information into sentences incrementally and immediately without waiting for the crucial moment when they are crystal clear where to integrate the information into. Such incremental and immediate processing has been tested, as garden-path effects, in readers' parsing of structurally ambiguous clauses (Garnsey et al., 1997; Ferreira & Clifton, 1986; Trueswell et al., 1994).

When incremental processing is temporally maximized during sentence comprehension, it reveals the aspects of expectation in processing (Altmann & Mirković, 2009; Elman, 1990; 2009). Under this aspect, words or phrases are processed anticipatorily, prior to their occurrence, in a way that more expected information is easier to be processed than less expected information. There are a huge number of studies in this vein, in behavioral studies (Roland et al., 2012; Yun & Hong, 2014), eye-tracking studies (Ashby et al., 2005; Ehrlich & Rayner, 1981; Frisson et al., 2005; Staub, 2011; 2015), and ERP studies (Aurnhammer et al., 2021; Bicknell et al., 2010; DeLong et al., 2005; Federmeier et al., 2007).

Well-known neurological evidence observed by DeLong et al. (2005), for instance, supported that a particular word for a given context was anticipated before it was encountered. The electronic amplitudes associated with less-expected sentence-final words (e.g., *airplane in The day was breezy so the boy went outside to fly a/an …*) went more negatively at the time course of 400 ms from the onset of the target words than the amplitudes associated with more-expected words (e.g., *kite* in *The day was breezy so the boy went outside to fly a/an* …). More crucially, the N400 emerged early from the position of the indefinite article *a* or *an* before the target nouns; for example, given the sentence fragment, *The day was breezy so the boy went outside to fly a/an* …, the N400 corresponding to *an*, compared to *a*, was significantly larger because readers could recognize instantly that the indefinite article, an, did not phonologically coordinate with the most-likely word, *kite*. Readers were extremely fast to compute *kite* as the most likely word choice, immediately after they processed the contextual information of *the day-breezy-the boy-go-outside-fly.*

The DeLong et al.'s (2005) results, including other similar findings, reveal that if a reader is a successful anticipatory processor, she is able to dynamically exploit the cues from contexts and is cognitively fast to compute potential upcoming lexical/structural choices that could appear next. In addition, she needs to own a sufficient size of working memory to manage all related information while she tries to integrate incoming information incrementally to the representations that are being constructed. Put it simply, although she is extremely busy with the linguistic task, she is cognitively efficient in completing the task successfully. However, are all readers cognitively efficient as good as they should be? Do all readers take similar strategies during sentence comprehension? Not much has been known to what degree and how systematically there would be individual differences in expectation-based sentence comprehension. In this study, we aimed to investigate, using a Korean dative construction, how readers' neurological responses would differ as a function of readers' working memory capacity with the view of expectation-based sentence comprehension.

1.1 Previous studies

More expected words or phrases are likely to be more strongly activated with less processing cost during comprehension (c.f., Hale, 2001; Levy, 2008; Roland et al., 2012). Studies in this field have suggested that the kinds of activated information associated with upcoming words or phrases are not limited to one specific information but various kinds of information at multiple linguistic levels such as syntactic information (e.g., Boston et al., 2008; Demberg et al., 2013), thematic role information (e.g., Yun & Hong, 2014), semantic featural information (e.g., Federmeier & Kutas, 1999; Roland et al., 2012), world-knowledge information (e.g., DeLong et al., 2005; Kutas & Federmeier, 2011), and so on.

For example, Altmann and Kamide (1999) tested that the immediate activation of verb's argument information has been anticipatorily used to regulate the semantic featural likelihood of following argument fillers. They observed anticipatory eye movements to a depicted object (e.g., cake of the [+edible] information), prior to the auditory onset of the word referring to the object, in the event encoded by a verb (e.g., the eating event encoded by the verb, eat). Similarly, Roland et al. (2012) found that readers felt easier to process words by the degree to which the words were semantically similar to the other possible word choices that could have occurred instead of the target words. For instance, for a given sentence fragment like *The man stabbed the old king with a(n) …*, unlikely instrument words like *machete* were relatively easy to process because *machete* shares similar semantic features (e.g., pointy or having a blade) with other possible choices like *knife* or *dagger* that could occur instead. In contrast, for a sentence fragment like *The man attacked the old king with a(n) …*, unlikely instrument words like *stone* might be relatively difficult to process because *stone* does not share much semantic featural information with other possible word choices like *knife, gun*, or *words.*

Note that in head-initial languages like English, early-introduced verbs, together with subjects, contribute to cue what is coming up next and constrain the semantic features of upcoming words. For such a sentence structure (i.e., SVO word order), most-predictable words are likely to be rapidly integrated upon the recognition of the words without readers' additional cognitive cost. In contrast, the integration of less-predictable words might be delayed with readers' extra cognitive efforts, primarily because readers should update their representation by shifting it from a predictable word choice to a less-predictable word choice.

The expectation-based sentence comprehension has been observed in verb-final languages including Korean and Japanese. Of interest, it has been claimed that the grammatical function carried by case-markers, used in those languages, played an important role in eliciting the expectation of upcoming arguments even before the occurrence of sentence-final main verbs (Kamide et al., 2003; Yun et al, 2017). Using a visual world paradigm in Japanese sentence comprehension, Kamide et al. (2003) observed anticipatory eye movements toward depicted objects corresponding to patient roles, at adverbial phrases before main verbs, in sentence fragments in which recipient roles with dative case markers were mentioned. However, they did not find such anticipatory eye movements toward depicted objects corresponding to recipient roles in sentence fragments in which patient roles with accusative case markers were mentioned. In a similar vein, Yun et al. (2017), through an eye-tracking reading study using Korean dative sentences,[1] observed that Korean readers took longer to read the phrases of

[1] The Korean dative constructions have two case-marking patterns comparable to English dative alternations: 1) Prepositional Object pattern (i.e., dative marker (*-eykey*) + accusative marker (*-(l)ul*) (or accusative marker (*-(l)ul*) + dative marker (*-eykey*)) and 2) Double Object pattern (i.e., accusative marker (*-(l)ul*) + accusative marker (*-(l)ul*)). Some linguistics (Lee, 2020; Yoon, 2015) accept the both patterns as Korean dative constructions, whereas other studies claimed that double accusative construction is less preferred that Korean speakers find it awkward and unacceptable (Cho & Jeon, 2015; Shin, 2020). In fact, the double object construction occurs so infrequently that it hardly generates any probability differences at all and if it does, the construction sounds very unnatural (c.f., Park & Yi, 2021). For

patients and recipients than those of recipients and patients. These studies revealed that Korean and Japanese readers were proficient at developing the expectation of upcoming patient roles when recipient roles were provided, but not vice versa, leading to the processing benefit of patient roles occurring after recipient roles rather than recipient roles occurring after patient roles. Crucially, the integration of the expected role information into a sentence fragment could be initiated even before the occurrence of main verbs.

More specifically, Yun and Hong (2014) have demonstrated a statistical probability model of predictive comprehension in processing Korean dative sentences, using the sentences like (1a-d). The reading times differences of target words (i.e., the third word written in bold) across conditions were straightforward. First, Role Predictability played a significant factor in predicting reading times variances, meaning that words whose thematic roles were highly predictable, as found in (1a-b) where themes (i.e., NP with *–ul*) followed recipients (i.e., NP with *–eykey*), were faster to read than words whose thematic roles were weakly expected, as found in (1c-d) where recipients (i.e., NP with *–eykey*) followed themes (i.e., NP with *–ul*), regardless of whether the words themselves were likely or unlikely for given contexts. Second, Word Predictability was a significant factor, indicating that likely words for given contexts, as in *sinpuncung-ul* (meaning ID card) in (1a), were faster to process than unlikely words for given contexts, as in *sinpuncung-ul* in (1b).

---

the current study, we consider only the prepositional object construction, with and without canonical word order, as a Korean dative construction.

(1a) Likely Role, Likely Word:
Chelwu-ka | kyengchal-eykey | **sinpwuncung-ul** | tangtanghakey | ceysi-hayss-ta
Chelwu-NOM policeman-DAT ID card-ACC proudly showed
Chelwu showed a policeman (his) ID card proudly.

(1b) Likely Role, Unlikely Word:
Minhoka | moteyl-eykey | **sinpwuncung-ul** | tangtanghakey | ceysi-hayss-ta
Minho-NOM model-DAT ID card-ACC proudly showed
Minho showed a model (his) ID card proudly.

(1c) Unlikely Role, Likely Word:
Hochel-ika | sinpwuncung-ul | **kyengchal-eykey** | tangtanghakey | ceysihayssta
Hochel-NOM ID card-ACC policeman-DAT proudly showed
Hochel showed (his) ID card to a policeman proudly.

(1d) Unlikely Role, Unlikely Word:
Wuseng-ika | sinpwuncung-ul | **moteyl-eykey** | tangtanghakey | ceysihayssta
Wuseng-NOM ID card-ACC model-DAT proudly showed
Wuseng showed (his) ID card to a model proudly.

Unlike English readers, Korean readers might go through different steps of sentence comprehension. For example, using the declarative sentences of SOV construction, as shown in (1a-d), Korean readers should exploit the case-marker information to generate which thematic role an upcoming word would take. They should actively use the contextual information to compute a range of possible word and structural choices for a given context. In addition, before they encounter a sentence-final verb, Korean readers should store previously-introduced arguments in their working memory to complete the integration of the argument fillers into the main verb. This is what Korean sentence comprehension is different from English sentence comprehension.

1.2 Individual differences in sentence processing

The benefit of pre-activation to process upcoming information, prior to its occurrence, largely depends on how actively readers can exploit available linguistic cues from contexts to pre-activate related information and how efficiently they are able to apply them anticipatorily in the processing of incoming words and phrases. In other words, readers' cognitive flexibility or storage to maintain and coordinate their linguistic knowledge should attribute to the goodness of expectation-based sentence comprehension.

In fact, readers' cognitive resources like Working Memory (WM) have been pointed out as a critical factor in accounting for the goodness of language processing. For example, WM, which refers to an individual's limited-capacity system responsible for short-term maintenance, storage, and manipulation of information, plays an essential role in the good quality of lexical representations (Bell & Perfetti, 1994; Hamilton et al., 2013; Perfetti, 2007; Perfetti et al., 2005) and on the encoding of speech in noise during language production (Akeroyd, 2008; Rönnberg et al., 2013). Overall, readers with high WM, compared to those with low WM, were faster to process sentences and resolved lexical and structural ambiguity more easily (Clifton et al., 2003; Daneman & Carpenter, 1980; Ericsson & Kintsch, 1995; Just & Carpenter, 1992; King & Just, 1991; Lewis & Vasishth, 2005; Matzke et al., 2002; Van Petten et al., 1997; Vos et al., 2001; Waters & Caplan, 1996).

There have been quantitative and qualitative differences observed in accounting for individual variances in sentence comprehension. Some studies have detected quantitative differences such that readers with high WM are faster and more robust in the use of lexical and syntactic information than those with low WM (Miyake et al., 1994; Van Petten et al., 1997; Vos et al., 2001). Other studies have demonstrated qualitative differences such that readers with low WM tend to focus more on using lexical information, whereas those with high WM are more sensitive to using combinatorial information (Nakano et al., 2010; Tanner, 2013). For example, high WM readers applied semantic information (e.g., animacy) to guide their initial parsing preferences in garden-path sentences (e.g., reduced relative clauses), but low WM readers seemed unable to use such information (Bornkessel et al., 2004; Just & Carpenter, 1992; Nakano et al., 2010).

Recently, Ryskin, Levy, and Fedorenko (2020) have proposed several hypotheses that could be used to test individual differences in expectation-based sentence comprehension: First, cognitive resources like WM are necessarily required for readers to maintain a faithful representation of the preceding linguistic context which in turn helps readers anticipate upcoming information for the given context or integrate diverse kinds of information into sentences (see Futrell & Levy, 2017). Second, WM is required for readers to generate a list of possible linguistic choices for an upcoming position and maintain them as actively as possible in the working memory space if multiple possible continuations are developed (Just & Carpenter, 1992; King & Just, 1991). Third, the success of expectation-driven linguistic prediction might represent the function of readers' inhibitory and selection mechanisms. That is, although it is important to keep the wide distribution of possible choices in their WM, readers might need to select the most-likely choice while inhibiting less-likely choices as efficiently as possible (Mirman et al., 2011; Nozari et al., 2016a; 2016b). In addition to Ryskin et al.'s hypothetical claim, we also propose that cognitive flexibility is required, especially when readers' expectation is not satisfied; when readers should shift their attention from a likely choice that they would expect to encounter to an unlikely choice that they would not expect to encounter. However, the role of readers' WM in expectation-based predictive sentence comprehension has not been tested systematically with actual online empirical data.

As an interim summary, a number of studies have demonstrated that readers' active use of lexical and structural information and event knowledge elicits anticipatory processing of yet-to-be-encountered information. Indeed, it is highly speculated that successful expectation-based processing might depend on the availability of readers' cognitive resources like WM [c.f., Ryskin et al.'s (2020) claim.] Yun and Hong (2014) suggested that Korean readers actively and anticipatorily used both argument information conveyed by case markers and contextual information to facilitate the integration of upcoming information during the processing of Korean dative sentences. Yet, they have not investigated whether Korean readers' expectation-based sentence comprehension would be observed regardless of readers' cognitive resources. In this study, we aimed to examine whether the goodness for expectation-driven comprehension might differ due to the availability of readers' cognitive resources such as WM. Using an ERP paradigm, we attempted to test whether the individual differences in readers' WM would result in significant variations in the predictive use of lexical and structural information during sentence comprehension.

The ERP components that we were interested in were the negativity peaking at the time window of 400 ms, known as the N400 effect, and the positivity peaking at the time window of 600 ms, known as the P600 effect. First, the N400 is a negative deflection peaking around 400 ms after a stimulus-onset, usually with a central-parietal distribution. The N400 occurs when semantically anomalous words are read, compared to semantically-normal words, (Kutas & Hillyard, 1980) and when semantically unexpected words, compared to semantically-expected words, are encountered (Federmeier & Kutas, 1999; Kutas & Hillyard, 1984; Lau et al., 2008). Second, the P600 is a positive deflection that starts at about 500 ms after a stimulus-onset and lasts several hundred milliseconds with central-parietal and sometimes frontal distributions. P600 amplitudes over central-parietal sites are increased in particular when readers encounter syntactic violations, syntactically unexpected words, or syntactically complex sentences (Hagoort et al., 1993; Kaan et al., 2000; Osterhout & Holcomb, 1992; Osterhout & Mobley, 1995). Slightly different from the P600 due to structural violation, the P600 effects, often called as semantic P600 effect, have been detected with sentences when implausible thematic role assignments were involved (Hoeks et al., 2004; Kim & Osterhout, 2005), when new discourse referents were introduced (Burkhardt, 2006; 2007), irony (Regel et al., 2011; Spotorno et al., 2013), or when other pragmatic factors had to be processed (Delogu et al., 2018; Hoeks et al., 2013). Finally, there is the anterior P600 which reflects readers' rapid shift from their pre-activated representation to an unexpected representation immediately after they encounter an unexpected word (Kuperberg et al., 2020). This P600 has been known to elicit when readers try to pay their special efforts for reintegration by updating their representation while suppressing incorrectly pre-activated information (Kuperberg et al., 2020).

For this study, we expected to observe that the N400 effect would occur when unexpected words or word phrases were encountered and that the P600 effect would elicit when unexpected words or word phrases were integrated. However, we did not make a concrete hypothesis about where (at which sites) the P600 effect would emerge. More crucially, the goal of this study was to investigate whether the N400 and the P600 elicited, as a sign for expectation-based sentence processing, would differ quantitatively or qualitatively by the size of readers' WM during the processing of Korean dative sentences.

2. Materials and Methods

2.1 Participants

Thirty-nine university students (27 female, 23.3 years old on average) in Seoul took part in the ERP study. All participants were right-handed and had no reading difficulties. They were all native Korean speakers and no one experienced living in foreign-language-speaking countries for more than three years. They received 30,000 won to compensate for their participation. Out of the 39 participants, the data from 9 participants had to be excluded due to unexpected technical failures and participants' negligent behaviors. The working memory capacity of the remaining 30 participants was measured by using Reading Span Task (Daneman & Carpenter, 1980). The scores were ranged from 1 corresponding to a low reading span to 5 corresponding to a high reading span. The 30 participants were grouped to high WM Group (N = 15, M = 4.23 SD = .56) and low WM Group (N = 15, M = 2.83, SD = .24).[2]

2.2 Materials

We used 40 sets of dative sentences that were partially taken from Yun and Hong's (2014) study. Each sentence consisted of five regions (i.e., from R1 to R5), as indicated in Table 1. The areas of interest for this study were the words at R3 (recipients or themes), R4 (sentence adverbs), and R5 (verbs). We kept the manipulation of the previous study: Role Predictability (RP, hereafter) indicating that an upcoming thematic role was strongly or weakly expected for given contexts in sentences and Word Predictability (WP, hereafter) indicating that an incoming word was likely or unlikely for a given context in each role condition.

First, words at R3 differed by two factors. One was the RP that the roles

[2] We used a medium-cut score to regroup 30 participants into two independent groups.

associated with incoming words (i.e., themes) were highly expected, as in (2a-b), where recipients following agents were presented, whereas the roles associated with incoming words (i.e., recipients) were unlikely, as in (2c-d), where themes following agents were presented. The other was the WP that an incoming word was likely for a given event, as in (2a)-(2c), or unlikely, as in (2b)-(2d). Second, words at R4 (i.e., adverbs) and words at R5 (i.e., verbs) were the same across the conditions. However, they differed in that they appeared in the high RP condition, as in (2a-b), in which themes followed recipients, whereas they appeared in the low RP condition, as in (2c-d), in which recipients followed themes.

Table 1. An example set of experimental materials

RP	WP	R1	R2	R3	R4	R5
(2a) High	High	*Chelswu-ka* Chelswu-NOM	*wuncensa-eykey* driver-DAT	*chapi-lul* fare-ACC	*emchengnakey* too much	*cipwulhayssta* paid
(2b) High	Low		*oyyaswu-eykey* player-DAT	*chapi-lul* fare-ACC		
(2c) Low	High		*chapi-lul* fare-ACC	*wuncensa-eykey* driver-DAT		
(2d) Low	Low		*chapi-lul* fare-ACC	*oyyaswu-eykey* player-DAT		

Note. RP refers to role predictability and WP does word predictability.

To generate experimental stimuli and norm the factors for our manipulation, we conducted a series of cloze tasks to measure the predictability of roles and words. As shown in Table 2, the RP was much higher in the high RP condition than in the low RP condition, meaning that themes NPs were dominantly produced at 73% for the sentence fragments of Recipient NPs following Agent NPs (e.g., *Chelswu-ka wuncensa-eykey* _____) but that recipient NPs were produced only at 3% for the sentence fragments of Theme NPs following Agent NPs (e.g., *Chelswu-ka chapbi-lul* _____). Similarly, the WP was higher in the high WP condition than in the low WP condition, meaning that specific likely NPs (e.g., *chapbi* meaning fare) were produced at 15% for given sentence fragments (e.g., *Chelswu-ka wuncensa-eykey* _____)[3] but specific unlikely NPs (*chapbi* meaning fare) were produced only at 0.3% for sentence fragments (e.g., *Chelswu-ka oyyaswu-eykey* _____).

Additionally, we checked whether experimental sentences were plausible enough by conducting a 7-scale rating task. The sentences from all conditions were rated at around point 5, meaning that experimental sentences sounded fairly natural. Finally, at R3 where neural responses associated with likely words were to be compared to those associated with unlikely words, it was important to control for lexical properties like lexical frequency between target words. The log-transformed lexical frequencies of target words in the high RP condition were the same when the words were likely (M = 2.91) and unlikely (M = 2.91). The log-lexical frequencies of target words in the low RP condition were not different when the words were likely (M = 3.0) and unlikely (M = 2.68). Because case markers attached to themes (i.e, *-l/ul*) were longer than those attached to recipients (i.e., *-ekey*), the length of words at R3 was systematically longer in the low RP condition than in the high RP condition. However, because we tested the effect of WP in each RP condition, respectively, we did not need to control for the word length differences.

[3] In computing word predictability, we used cloze completions almost as raw as they were produced, without any edit, instead of trying to combine similar completions. Thus, although the word probability for likely words could sound relatively low, the selected words were fairly predictable for given contexts.

Table 2. Means (standard deviations) of RP and WP across conditions

Condition	P(Role)	P(Word)	Sentence plausibility
High RP, High WP (e.g., 2a)	.73 (.10)	.15 (.07)	5.76 (.75)
High RP, Low WP (e.g., 2b)		.003 (.00)	5.4 (1.02)
Low RP, High WP (e.g., 2c)	.03 (.0)	.17 (.08)	5.41 (.89)
Low RP, Low WP (e.g., 2d)		.001 (.00)	4.86 (.97)

The 160 experimental sentences (40 sentences x 4 conditions) were intermixed with 200 filler sentences. The half of fillers were non-sensical, whereas the other half were sensical. The experimental set of 360 sentences was divided into 4 session blocks.[4] The presentation of the session blocks was counterbalanced across four presentation lists. The participants were randomly assigned to one of the four lists.

2.3 Procedure

Participants were seated at a distance of 70 cm (27.55″) from the center of the 19″ display monitor. Before starting the experiment, instructions were presented on the screen, and the electrodes were attached to the participants' heads during this time. Sentences were presented in the center of the screen using a Rapid Serial Visual Presentation method. Each sentence

[4] For an ERP study, having enough experimental items per condition is important in order to obtain accurate data (Luck. 2005). Because there are many cases where it is not easy to generate more than 36-40 sentences per condition, studies often solve this issue by opting for a block design. Thus, all participants are exposed to the same item under the same condition with different presentation order of experimental items. In addition, given that ERPs are sensitive to lexical properties, a compounding effect elicited by using various lexical items could occur. For this reason, using a block design is often favored in controlling for any unexpected lexical effect.

began with a “+” sign lasting for 500 ms, and after a 200 ms interval, each region was presented for 700 ms followed by a 300 ms blank. At the end of filler sentences, a sensicality judgement question appeared. Participants were asked to judge whether or not the presented sentence would make sense grammatically, semantically or pragmatically. They were required to press a button as quickly as possible. J and F keyboards were used as an indicator of sensicality, and the two keyboards were set differently for each participant. All experimental sentences were randomly presented so that the effect of the presentation order of target sentences was prevented as much as possible. Before starting the experiment, participants were allowed to adapt naturally to the experiment using eight practice sentences. The experiment lasted almost 90 minutes and participants were allowed to rest between the blocks.

EEG was recorded using a BrainAmp direct current (DC) amplifier (Brain Products, Germany) with 32 Ag/Agcl electrodes in an actiCAP (Brain Products, Germany). An electrode attached to the tip of the nose was used as the reference and FCz as the ground channel during the recording. The rest 30 electrodes (except reference and ground channels) were located corresponding to the International 10–20 system. Mastoids channels were re-referenced during the pre-processing of the ERP analysis. An independent component analysis (ICA) method (Makeig et al., 1997) was applied for the ocular correction. The impedance of all electrodes was kept below 5 kΩ prior to data recording. The raw EEG data were recorded at 250 Hz with a high-pass filter (0.1 Hz), and an offline band-pass filter (0.1–30 Hz) was applied in the pre-processing to display the ERP components clearly.

2.4 Analysis

For the behavioral analysis, we computed the accuracy rates of the participants’ judgments. As for the 30 participants, they made correct judgments at 80.86% on average. No participants whose accuracy rates were below 80% were included for further analysis because those participants were likely to process sentences as sincerely as we put a strong trust on their responses. For the ERP analysis, we segmented each epoch with 900 ms from each stimulus onset time; the segments were aligned to a 100 ms pre-stimulus baseline. Artifacts surpassing an amplitude of ± 100 μV, or higher than 50 μV within a moving 4 ms interval, were excluded from further processing. This led to the data removal approximately at 2.5%. Considering the typical N400 and P600 time windows and the visual inspection of grand-averaged brain waveforms and topographic maps, 300–400 ms, 400–500 ms, 600–700 ms, and 700–800 ms were used for the predictability effect elicited at the word of region 3, 4, and 5.[5] Electrodes were grouped into midline and lateral ROIs separately: Fz, Cz, Pz, and Oz channels were used as ROIs for midline analyses and the average value of the following six channels were used as lateral ROIs: left-anterior (LA): FC1, FC5, and F3; left medial (LM): C3, CP1, and CP5; left-posterior (LP): P3, P7, and O1; right-anterior (RA): FC2, FC6, and F4; right medial (RM): C4, CP2, and CP6; and right-posterior (RP): P4, P8, and O2. We conducted repeated measures ANOVAs and paired t-tests on grand-averaged voltages of all the separate comparisons to detect subtle differences between conditions at different ROIs. The Greenhouse-Geisser correction was applied when effects with more than one degree of freedom were evaluated in all the

[5]As for many ERP studies using English, it is common to detect the N400 in the window of 300-500 ms (Kutas & Hillyard, 1980) and the P600 in the window of 600-700 ms (or 800 ms) (Osterhout & Holcomb, 1992). However, the detection window for the N400 slightly differs across languages (Moreno & Kutas, 2005), such that the N400 is detected earlier for Korean sentences than for English ones (Yoon, 2008). In addition, if an experiment aims to capture newly-occurring potential effects instead of replicating already-found (or predicted) effects, detailed time-window analyses could be conducted through visual inspection of averaged-ERP components (Neufeld et al, 2016). Thus, for the proper reflection of potential effects that may vary depending on a language and a time window, we carried out detailed time-window analyses at each unit of 100 ms.

overall and separate statistical comparisons.

3. Results

3.1 Results at R3

Figure 1 illustrates the results elicited in the processing of words at R3 across conditions. We reported statistical results at the time window in 300-400 ms and 400-500 ms. No significant results were observed at the other time windows of 600-700 ms and 700-800 ms.

300-400 ms: We observed significant a three-way interaction (Role Predictability x Word Predictability x WM), F(1, 28) = 5.95, p = .02; at Midline (F(1, 28) = 5.83, p = .02), LA (F(1, 28) = 5.83, p = .02), RA (F(1, 28) = 13.32, p = .00), and RM (F(1, 28) = 8.60, p = .01). The main effect of WP was observed significantly at LA (F(1, 28) = 4.56, p = .04); RA (F(1, 28) = 4.23, p = .05), and marginally significantly at Midline (F(1, 28) = 3.78, p = .06); RM (F(1, 28) = 3.88, p = .06). To understand the patterns of the three-way interaction, we separated the data by RP (high RP condition and low RP condition) and WM groups (high WM and low WM), and tested the effect of WP in each separate group. First, as for the high WM group, no significant differences were observed between likely word and unlikely word at all ROIs when thematic roles associated with target words were highly expected (i.e., high RP condition). In contrast, in the condition where thematic roles were not highly expected (i.e., low RP condition), the amplitudes associated with unlikely words went more negatively than those associated with likely words. The significant N400 elicited at Midline (t(14) = 2.27, p = .04), LA (t(14) = 2.82, p = .01), LP (t(14) = 2.12, p = .05), RA (t(14) = 2.69, p = .02), RM (t(14) = 2.55, p = .02), and RP (t(14) = 2.27, p = .04), and marginally significant at LM (t(14) = 1.91, p = .08). Second, as for the low WM group, the significant N400 between unlikely words and likely words in the high RP condition appeared at RA (t(14) = 2.58, p = .02), and

RM (t(14) = 2.77, p = .02), and marginally significant at Midline (t(14) = 2.04, p = .06). In the low RP condition, the N400 was also significant at RA (t(14) = -2.12, p = .05) and marginally significant at RM (t(14) = -1.88, p = .08).

400-500 ms: We observed no significant three-way interactions (Role Predictability x Word Predictability x WM) at all ROIs. The main effect of WP was observed significantly at RA (F(1, 28) = 7.11, p = .01) and marginally significantly at Midline (F(1, 28) = 3.56, p = .07). Since we aimed to test the effect of WP across different WM groups, we conducted further analyses. The N400 between unlikely words and likely words in the low RP condition occurred only from the low WM group at LA (t(14) = 2.34, p = . 04) and LM (t(14) = 2.13, p = .05).

In summary, as we hypothesized, the N400 between likely and unlikely words emerged at the position of the third word (i.e., R3) where thematic roles associated with target words were highly predictable (i.e., high RP condition) or less predictable (i.e., low RP condition), as found in many previous studies (e.g., Federmeier & Kutas, 1999; Kutas & Hillyard, 1984; Lau et al., 2008). Crucially, we found quantitatively different patterns between the high WM readers and the low WM readers. That is, high WM readers showed the N400 only in the condition where thematic roles associated with target words were not highly expected (i.e., low RP condition) in 300-400 ms post-stimulus onset. They did not elicit significant negative responses to unlikely words of which thematic roles were highly expected (i.e., high RP condition). However, low WM readers showed the N400 both when thematic role expectation was high (i.e., high RP condition) in 300-400 ms post-stimulus onset and when it was low (i.e., low RP condition) in 300-400 ms post-stimulus onset and 400-500 ms post-stimulus onset. The key results at R3 are summarized in Table 3. These results indicated that readers with high WM capacity were fast in using upcoming lexical information (c.f., Van Petten et al., 1997) and insusceptible to unlikely lexical information as long as thematic role information associated with words was strongly guaranteed (that is, highly expected).

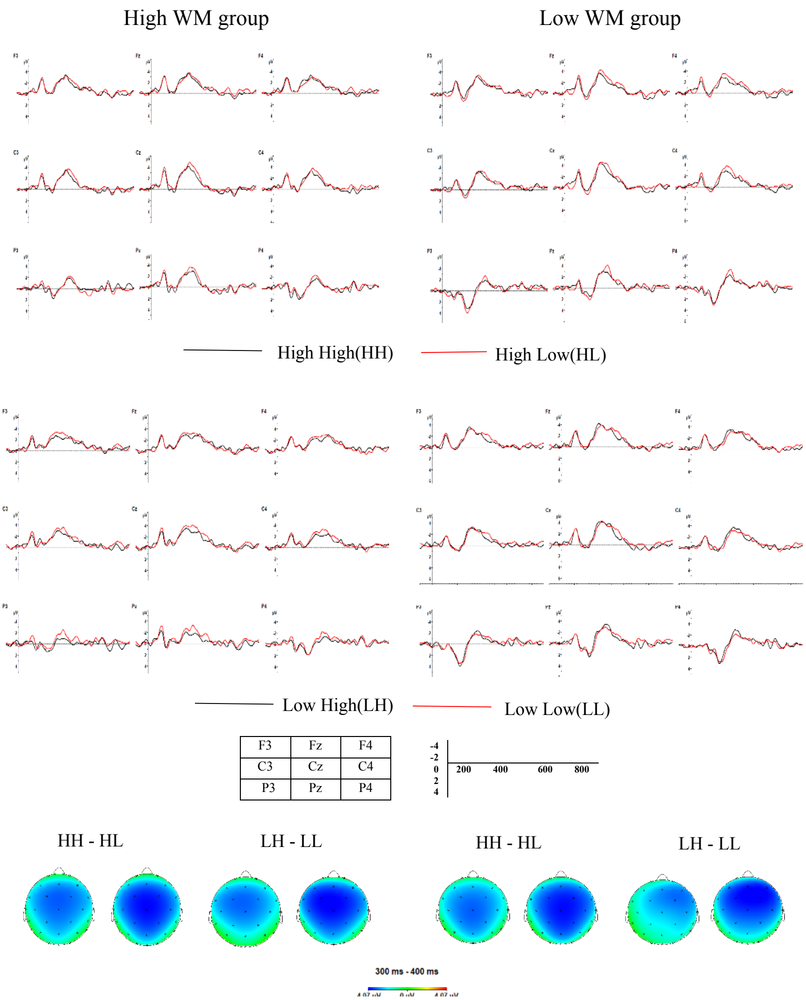

Figure 1. ERPs at all the ROIs at R3 (HH referring to high RP and high WP; HL referring to high RP and low WP; LH referring to low RP and high WP; LL referring to low RP and low WP), and scalp topographic distributions of different conditions by high and low WM groups. Figures for the high WM group are illustrated on the left and those for the low WM group are on the right.

3.2 Results at R4

Figure 2 illustrates the results elicited in the processing of words (i.e., sentential adverbs) at R4 between the high RP condition and the low RP condition. Recall that all three arguments (i.e., agents, recipients, and themes) were presented before R4 in an agent-recipient-theme order in the high RP condition and in an agent-theme-recipient order in the low RP condition. Thus, in encountering the sentential adverbs at R4, readers were holding these arguments in their working memory, while they were trying to integrate the arguments into sentences. Since the same adverbs were used across the conditions, there was no differences in terms of the lexical information. We reported statistical results in the time window of 600-700 ms post onset of adverbs. No effects were observed in the other time windows.

600-700 ms: We observed significant two-way interactions (Role Predictability x WM): $F(1, 28) = 6.52$, $p = .02$ across all ROIs; in particular, at Midlines ($F(1, 28) = 4.98$, $p = .03$), LA ($F(1, 28) = 12.60$, $p = .00$), RA ($F(1, 28) = 9.47$, $p = .01$), LM ($F(1, 28) = 6.45$, $p = .02$), and RM ($F(1, 28) = 5.85$, $p = .02$). These two-way interactions were further examined by testing the effect of RP in each WM group. First, as for the high WM group, the P600 effect was significant at LA ($t(14) = -2.58$, $p = .02$ and RA ($t(14) = -2.40$, $p = .03$). Second, from the low WM group, the significant negativity emerged at Midline ($t(14) = 2.63$, $p = .02$), LA ($t(14) = 2.44$, $p = .03$), LM($t(14) = 2.88$, $p = .01$), and RM ($t(14) = 2.86$, $p = .01$), but marginally significant at RA ($t(14) = 2.0$, $p = .06$).

In short, we found qualitatively different ERP patterns between high WM readers and low WM readers at the adverbs (i.e., R4). From the high WM readers, we observed the P600 effect with anterior distributions when the targets were encountered after an unexpected order of role clusters (i.e., low RP condition) than after an expected order of role combinations (i.e., high RP condition). The positivity with anterior distributions suggested that high

WM readers might have been involved in the process of reintegration with special efforts (cf., Kaan et al., 2000; Kann & Swaab, 2003) by updating their representations while they were suppressing incorrectly pre-activated information (Kuperberg et al., 2020). The high WM readers might expect to encounter main verbs where they could complete the integration of the arguments. However, low WM readers showed the negativity deflections at the same time window, suggesting that they had somewhat difficulty in maintaining the currently presented information. We discuss the slow negative waves with frontal distribution in the discussion section. In a nutshell, the response dichotomy at the adverbial region between the two WM groups hinted that only high WM readers, not low WM readers, actively engaged in the expectation-driven combinatorial processes in integrating arguments into sentences (c.f., Nakano et al., 2010). Also, see the key summary of ERPs at R4, illustrated in Table 3.

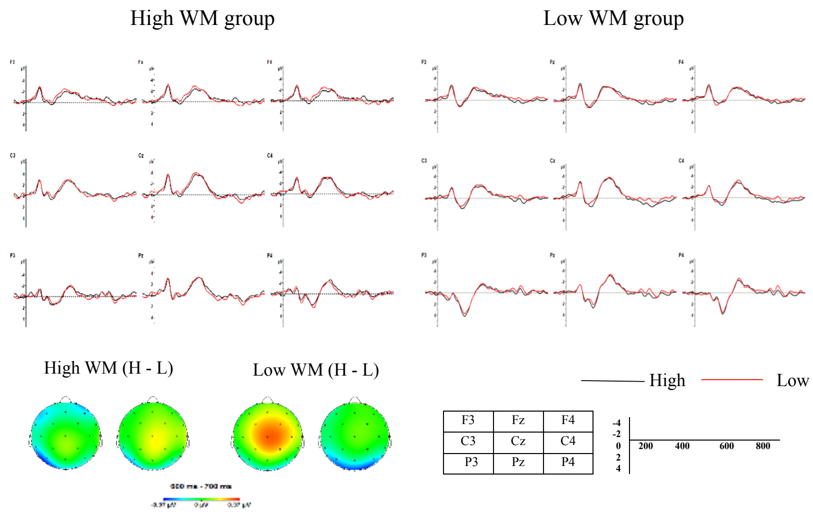

Figure 2. ERPs at all the ROIs at R4 (black: high RP condition, red: low RP condition), and scalp topographic distributions of different conditions by high and low WM groups

3.3 Results at R5

Figure 3 illustrates the results elicited in the processing of sentence-final main verbs at R5 between the high RP condition and the low RP condition. The same verbs were used across the conditions, and thus there was no difference in terms of the lexical information. Recall that all three arguments (i.e., agents, recipients, and themes) plus with sentential adverbs were presented before R5 in an agent-recipient-theme order in the high RP condition and in an agent-theme-recipient order in the low RP condition. In encountering main verbs at R5, readers needed to retrieve previously-presented arguments for the successful completion of argument integration. We reported statistical results at the time window of 300-400 ms and 700-800 ms. No effects were observed in the other time windows.

300-400 ms: We observed significant two-way interactions (Role Predictability x WM): $F(1, 28) = 6.33$, $p = .02$; in particular, significant at LP ($F(1, 28) = 4.33$, $p = .05$), and marginally significant at LM ($F(1, 28) = 3.70$, $p = .07$). There was no interesting differences observed for the high WM group at all ROIs. However, as for the low WM group, amplitudes for target verbs went more negatively in the low RP condition than in the high RP condition at midline ($t(14) = 3.18$, $p = .01$), LM ($t(14) = 3.37$, $p = .01$), LP ($t(14) = 3.86$, $p = .01$), RM ($t(14) = 2.79$, $p = .02$), and RP ($t(14) = 2.34$, $p = .04$).

700-800 ms: Only the low WM group elicited the late negativity effect at Midline ($t(14) = 2.24$, $p = .04$), RM ($t(14) = 2.53$, $p = .02$), RP ($t(14) = 2.34$, $p = .04$), and marginally significant at RA ($t(14) = 2.02$, $p = .06$).

To be brief, in the processing of sentence-final verbs where all arguments had to be integrated into sentences, we observed qualitative differences between the two groups. The high WM readers did not yield particular deflections at verbs. In contrast, low WM readers were characterized with two significant results. First, the N400 reappeared, meaning that the ERP deflections in response to verbs in the high RP condition moved more negatively than those with in the low RP condition. This result suggested that the low WM readers were involved in lexical processing as if they would encounter an unexpected word. Second, the late negativity, known as LPN, occurred at 700-800 ms post-stimulus onset. The deflections aligned with verbs were more negatively-going in the processing of target verbs occurring in the low RP condition than in the high RP condition. This late negativity, known as a component referring to information retrieval (Meckinger et al., 2007), was not what we hypothesized in the beginning. For now, we thought that the disadvantages of low WM capacity might result in the difficulty in reconstructing previously provided episodic information while they were retrieving argument information and integrating it into contexts. We discuss more in detail later. Table 3 indicates the key summary of ERPs at R5.

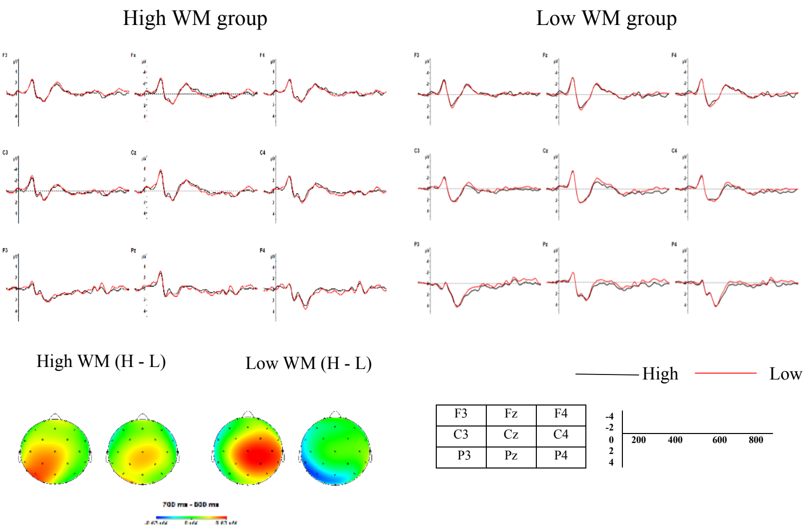

Figure 3. ERPs at all the ROIs at R5 (black: high RP condition, red: low RP condition), and scalp topographic distributions of different conditions by WM groups

Table 3. Key summary of ERP results across regions and conditions

		R3 *Chapi-lul (H-L)/ wuncensa-eykey (H-L)* (fare-drive)	R4 *emchengnakey* (too much)	R5 *cipwulhayssta* (paid)
High WM readers	High RP (canonical order)	None	**P600** (LA, RA at 600-700 ms)	None
	Low RP (noncanonical order)	**N400** (Midline, LA, LM, RA, RM, RP at 300-400 ms)		
Low WM readers	High RP (canonical order)	**N400** (Midline, RA, RM at 300-400 ms)	**Negativity** (Midline, LA, LM, RA, RM at 600-700 ms)	**N400** (Midline, LM, LP, RM, RP at 300-400 ms) **Late Negativity** (Midline, RA, RM, RP at 700-800 ms)
	Low RP (noncanonical order)	**N400** (RA, RM at 300-400 ms; LA, LM at 400-500 ms)		

4. Discussion

According to the perspective of expectation-based sentence comprehension, readers' active use of given lexical-structural information and event knowledge elicits anticipatory processing of yet-to-be-encountered information and facilitates the integration of incoming information. In this study, we aimed to examine how the goodness for expectation-driven comprehension systematically differs by readers' cognitive resources such as WM. Using an ERP paradigm, we demonstrated that the differences in readers' WM capacity led to crucial variations, quantitatively and qualitatively, on the anticipatory use of lexical and structural information in the processing of Korean dative sentences. Of our great interest, the asymmetrical distributions of the negativity and the positivity induced by readers, attributable to readers' WM differences, suggested that readers with low WM were more involved in lexical retrieval, whereas those with high WM were much more attentive to structural or semantic re-integration process. We argue that our results suggest that the lack of WM capacity could make readers fall behind in using lexical and structural information during sentence processing, in particular, for argument integration.

Our main findings are as follows. First, we observed that the negative deflections to unlikely words were higher than those to likely words. From the readers with high WM, the N400 effects occurred only when thematic roles associated with target words were unexpected, that is, when recipients followed themes noncanonically. No N400 in the expected role condition suggested that readers with high WM may not have significant difficulty in retrieving target words, although the words were not likely, as long as they were fairly sure that any word referring to a theme role would occur. However, this is not the case for readers with low WM. They revealed the N400, regardless of the presentation order of thematic roles, both when themes followed recipients and vice versa. The readers with low WM might be busy with updating new lexical information and have difficulty in lexical retrieval if the words were not likely. The quantitative differences of the N400 between the high WM group and the low WM group that we observed were in line with many previous studies showing individual differences (Miyake et al., 1994; Van Petten et al., 1997; Vos et al., 2001)

Second, the other quantitative differences between the two groups appeared in the window of 700-800 ms post-stimuli onset of sentence-final verbs. This late negativity in response to verbs occurred to both groups of readers (see the ERP deflections Figure 3). However, the responses from high WM readers did not reach a significant level of differences. As mentioned earlier, the late negativity, mostly with posterior distributions, has been known as an ERP component responsible for reconstructing prior episodes (Mecklinger et al., 2016) and contextual retrieval (Herron, 2007; Johansson & Mecklinger, 2003). Given the head-final structures in Korean, previously-presented arguments have to be retrieved at sentence-final verbs where arguments need to be fully integrated into the events encoded by verbs. Under this scenario, Korean readers might be somewhat suffering from heavy memory loads at sentence-final verbs, because they should retrieve (or maintain) previously-encountered arguments to complete the integration of arguments into sentences while they are required to process verbs' encoding information, simultaneously. Our results led to reasonable speculation such that low WM readers, in comparison to high WM readers, seem to have much severer difficulty in retrieving arguments or maintaining them in their WM for the complete integration of arguments.

Third, our results revealed interesting qualitative differences in the processing of adverbs between the two WM groups. The readers with high WM evoked the positivity in the time window of 600 post-stimuli onset of target words with anterior distributions (i.e., anterior P600), whereas the readers with low WM induced the negativity at the same time window with midlines and left/right anterior and medial distributions (i.e., N600). Note that the same adverbs were used across the condition, thus any neurological differences should not be attributed to lexical differences in target adverbs. Unlike the posterior P600 which is related to semantic integration (Deloug et al., 2019), the anterior P600 has been known as an ERP component that reflects readers' high-level shift from their preactivated situational model to a new unexpected situational model, immediately after they encounter a new unexpected incoming word (Kuperberg et al., 2020; also see Brothers et al., 2015; DeLong et al., 2014; Federmeier et al., 2010). Then, for our study, observing the anterior P600 means that readers with high WM might be involved in this shifting process when they encountered unexpected adverbs instead of expected verbs, and that the structural shifting was much harder after they processed arguments in an unexpected order (i.e., recipients following themes) than in an expected order (i.e., themes following recipients). Thus, to reach an updated interpretation, the difficulty that high WM readers were faced with, observed as the anterior P600, was more likely to be related to structural integration rather than lexical integration.

Unlike high WM readers, low WM readers did not seem to try integrating adverbs into sentences. Instead, they evoked the N600 in response to adverbs with anterior distributions, which simply reflected readers' heavy working memory load (see King & Kutas, 1995), indicating that readers felt a much heavier memory load in updating the information from the adverbs when arguments were presented in an unexpected order than in an expected order. This task-general N600, not specifically dependent on so-far established semantic representations, was found in some other studies (Cummings et al., 2006), as an indicator of working memory or general cognitive processes such as attention (Itoh et al., 2005; Koelsch et al., 2003; King & Kutas, 1995). In a nutshell, the qualitative differences of ERPs, observed at the adverbs, between the two WM groups indicated that high WM readers yielded a neurological sign that they were somewhat involved in the structural integration of early-presented arguments into sentences, whereas low WM readers did not reveal such a neurological evidence.

Finally, only readers with low WM induced the negative-going deflections in response to sentence-final verbs in the time window of 300-400 ms post stimulus onset (i.e., N400 effect), suggesting that readers might be surprised to encounter verbs more in the theme-recipient condition (i.e., low RP condition) than in the recipient-theme condition (i.e., high RP condition). This negativity effect reflects readers' difficulty in lexical retrieval, rather than semantic integration (see Deloug et al., 2019). Recall that the same verbs were used across the conditions, and thus the N400 was not likely to emerge due to any lexical differences. Nonetheless, the lexical difficulty associated with this N400 might not be entirely independent of the processing difficulty with structurally irregular, thus unexpected, sentences. Further fine-grained studies are needed for this matter.

Overall, our results show that low WM readers' responses were more biased to N400-dominant effects, namely lexical retrieval, whereas readers with high WM additionally generated a P600-dominant effect, namely semantic or structural integration. We observed the asymmetrical distribution of the negativity and the positivity between the two WM groups when they were involved in online sentence comprehension of Korean dative structures. In fact, we are not the only study that has shown the sort-of trade-off patterns between N400 and P600. Kim, Oines, and Miyake (2018) also reported that a larger verbal WM capacity was significantly correlated with a larger P600 and smaller N400 effect across individual readers. For a clearer picture of this matter, there need to be further studies that take into account other cognitive resources like cognitive control while extending target sentences to different types of linguistic constructions. At least for now, taken together with our results, we argued that the lack of WM capacity could make readers fall behind in the use of lexical and structural information during sentence processing, in particular, for argument integration.

Acknowledgements

This work was supported by the research grant of Jeju National University in 2018.

References

Akeroyd M. A. (2008). Are individual differences in speech reception related to individual differences in cognitive ability? *International Journal Audiology*, 47(2), S53-S71.

Altmann, G. T. M., & Kamide, Yuki. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, 73, 247–264.

Altmann, G. T. M., & Mirković, J. (2009). Incrementality and prediction in human sentence processing. *Cognitive Science*, 33, 583-609.

Ashby, J., Rayner, K., & Clifton, C., Jr. (2005). Eye movements of highly skilled and average readers: Differential effects of frequency and predictability. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, 58A(6), 1065–1086.

Aurnhammer, C., Delogu F, Schulz, M., Brouwer, H., & Crocker, M. W. (2021). Retrieval (N400) and integration (P600) in expectation-based comprehension. *PLoS ONE*, 16(9), e0257430. https://doi.org/10.1371/journal.pone.0257430

Balota, D. A., Pollatsek, A., & Rayner, K. (1985). The interaction of contextual constraints and parafoveal visual information in reading. *Cognitive Psychology*, 17(3), 364-390.

Bell, L. C., & Perfetti, C. A. (1994). Reading skill: some adult comparisons. *Journal of Educational Psychology*, 86, 244–255. doi: 10.1037/0022-0663.86.2.244

Bicknell, K., Elman, J. L., Hare, M., McRae, K., & Kutas, M. (2010). Effects of event knowledge in processing verbal arguments. *Journal of Memory and Language*, 63(4), 489-505.

Bornkessel, I. D., Fiebach, C. J., & Friederici, A. D. (2004). On the cost of syntactic ambiguity in human language comprehension: An individual differences approach. *Cognitive Brain Research*, 21, 11–21. http://dx.doi org/10.1016/j.cogbrainres.2004.05.007

Boston, M. F., Hale, J., Kliegl, R., Patil, U., & Vasishth, S. (2008). Parsing costs as predictors of reading difficulty: An evaluation using the Potsdam Sentence Corpus. *Journal of Eye Movement Research*, 2(1), 1, 1–12.

Brothers, T., Swaab, T. Y., & Traxler, M. J. (2015). Effects of prediction and contextual support on lexical processing: Prediction takes precedence. *Cognition*, 136, 135-149.

Burkhardt, P. (2006). Inferential bridging relations reveal distinct neural mechanisms: Evidence from event-related brain potentials. *Brain and Language*, 98(2), 159-168.

Burkhardt, P. (2007). The P600 reflects cost of new information in discourse memory. *Neuroreport*, 18(17), 1851-1854.

Clifton Jr, C., Traxler, M. J., Mohamed, M. T., Williams, R. S., Morris, R. K., & Rayner, K. (2003). The use of thematic role information in parsing: Syntactic processing autonomy revisited. *Journal of Memory and Language*, 49(3), 317-334.

Cummings, A., Čeponienėa, R., Koyamaa, A., Saygin, A., Townsend, J., & Dick, F. (2006). Auditory semantic networks for words and natural sounds. *Brain Research*, 1115, 92–107.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior,* 19, 450–466. http://dx.doi.org/10.1016/S0022-5371(80)90312-6

Delogu, F., Drenhaus, H., & Crocker, M. W. (2018). On the predictability of event boundaries in discourse: An ERP investigation. *Memory & Cognition*, 46(2), 315-325.

DeLong, K. A., Troyer, M., & Kutas, M. (2014). Pre-processing in sentence comprehension: Sensitivity to likely upcoming meaning and structure. *Language and Linguistics Compass*, 8(12), 631-645.

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8, 1117–1121.

Demberg, V., & Keller, F. (2008). Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition*, 109(2), 193-210.

Demberg, V., Keller, F., & Koller, A. (2013). Incremental, predictive parsing with psycholinguistically motivated tree-adjoining grammar. *Computational Linguistics*, 39(4), 1025-1066.

Elman, J. L. (1990). Finding Structure in Time. *Cognitive Science*, 14, 179–211.

Elman, J. L. (2009). On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. *Cognitive Science*, 33, 547–582.

Elman, J.L. (2012). Lexical knowledge with a lexicon?. *The Mental Lexicon*, 60, 1–33.

Ehrlich, S. F., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior*, 20, 641–655.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, 102, 211–245. http://dx.doi.org/10.1037/0033-295X.102.2.211

Federmeier K. (2007). Thinking ahead: the role and roots of prediction in language comprehension. *Psychophysiology*, 44(4), 491–505. doi:10.1111/j.1469-8986.2007.00531.x

Federmeier K., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, 41(4), 469–495. doi:10.1006/Jmla.1999.2660

Federmeier K., Kutas M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain Language*, 115(3), 149–161. doi:10.1016/j.bandl.2010.07.006

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research*, 1146, 75-84.

Ferreira, F., & Clifton Jr, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, 25(3), 348-368.

Ferreira, F., & Patson, N. D. (2007). The 'good enough'approach to language comprehension. *Language and Linguistics Compass,* 1(1-2), 71-83.

Ferreira, F., Christianson, K., & Hollingworth, A. (2001). Misinterpretations of garden-path sentences: Implications for models of sentence processing and reanalysis. *Journal of Psycholinguistic Research*, 30(1), 3-20.

Frank, S. L., & Bod, R. (2011). Insensitivity of the human sentence-processing system to hierarchical structure. *Psychological Science*, 22(6), 829-834.

Friederici, A.D. et al. (1998). First-pass versus second-pass parsing processes in a Wernicke's and a Broca's aphasic: electro-physiological evidence for a double dissociation. *Brain and Language*, 62, 311–341.

Frisson, S., Rayner, K., & Pickering, M. (2005). Effects of contextual predictability and transitional probability on eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(5), 862–877.

Futrell, R., & Levy, R. (2017, April). Noisy-context surprisal as a human sentence processing cost model. *In Proceedings of the 15th conference of the European chapter of the Association for Computational Linguistics*: Volume 1, long papers (pp. 688-698).

Garnsey, S. M., Pearlmutter, N. J., Myers, E., & Lotocky, M. A. (1997). The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language*, 37(1), 58-93.

Hagoort P., Baggio G., & Willems, R. (2009). Semantic unification In Gazzaniga MS (Ed.), *The Cognitive Neurosciences*, 4th Ed (pp. 819–836). Cambridge, MA: MIT Press.

Hagoort, P., Brown, C., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Language and Cognitive Processes*, 8(4), 439–483. doi:10.1080/01690969308407585

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. In Proceedings of the *Second Meeting of the North American Chapter of the Association for Computational Linguistics* (pp. 1–8). Pittsburgh, PA: Carnegie Mellon University.

Hamilton, S. T., Freed, E. M., & Long, D. L. (2013). Modeling reader and text interactions during narrative comprehension: a test of the lexical quality hypothesis. *Discourse Process*, 50, 139–163. doi: 10.1080/0163853X.2012.742001

Herron, J. E. (2007). Decomposition of the ERP late posterior negativity: Effects of retrieval and response fluency. *Psychophysiology*, 44(2), 233-244.

Hoeks, J. C., Stowe, L. A., & Doedens, G. (2004). Seeing words in context: the interaction of lexical and sentence level information during reading. *Cognitive Brain Research*, 19(1), 59-73.

Hoeks, J. C., Stowe, L. A., Hendriks, P., & Brouwer, H. (2013). Questions left unanswered: How the brain responds to missing information. *PloS ONE*, 8(10), e73594.

Itoh, K., Suwazono, S., Arao, H., Miyazaki, K. I., & Nakada, T. (2005). Electrophysiological correlates of absolute pitch and relative pitch. *Cerebral Cortex*, 15(6), 760-769.

Johansson, M., & Mecklinger, A. (2003). The late posterior negativity in ERP studies of episodic memory: action monitoring and retrieval of attribute conjunctions. *Biological Psychology*, 64(1-2), 91-117.

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

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, 15(2), 159-201.

Kim, A., Oines, L, & Miyake, A. (2018). Individual Differences in Verbal Working Memory Underlie a Tradeoff Between Semantic and Structural Processing

Difficulty During Language Comprehension: An ERP Investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44 (3), 406 – 420.

Kim, A., & Osterhout, L. (2005). The independence of combinatory semantic processing: Evidence from event-related potentials. *Journal of Memory and Language*, 52(2), 205-225.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language*, 30(5), 580-602.

King, J., & Kutas, M. (1995). Who Did What and When? Using Word- and Clause-Level ERPs to Monitor Working Memory Usage in Reading. *Journal of Cognitive Neuroscience*, 7(3), 376-395.

Koelsch, S., Grossmann, T., Gunter, T. C., Hahne, A., Schröger, E., & Friederici, A. D. (2003). Children processing music: electric brain responses reveal musical competence and gender differences. Journal of *Cognitive Neuroscience*, 15(5), 683-693.

Koelsch, S., Maess, B., Grossmann, T., & Friederici, A. D. (2003). Electric brain responses reveal gender differences in music processing. *Neuroreport*, 14(5), 709-713.

Kuperberg, G. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, 1146, 23–49.

Kuperberg, G., Brothers, T., & Wlotko, E. (2020). A Tale of Two Positivities and the N400: Distinct neural signatures are evoked by confirmed and violated predictions at different levels of representation. *Journal of Cognitive Neuroscience*, 32(1). 12–35. doi:10.1162/jocn_a_01465.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, 62, 621-647.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307, 161–163. http://dx.doi.org/10.1038/307161a0

Kutas, M., & Hillyard, S. A. (1980). Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biological Psychology*, 11(2), 99-116.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics:(de) constructing the N400. *Nature Reviews Neuroscience*, 9(12), 920-933.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106, 1126–1177.

Lewis, R. L., & Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science*, 29(3), 375-419.

Luck, S. J. (2005). *An Introduction to the Event-related Potential Technique.* Cambridge, Mass: MIT Press.

MacDonald, M., Spivey-Knowlton, M. J., Trueswell, J. C., & Tanenhaus, M. K. (1993). Context effects in syntactic ambiguity resolution: Discourse and semantic influences in parsing reduced relative clauses. *Canadian Journal of Experimental Psychology/Revue canadienne de Psychologie Expérimentale,* 47(2), 276.

Matzke, M., Mai, H., Nager, W., Rüsseler, J., & Münte, T. (2002). The costs of freedom: An ERP-study of non-canonical sentences. *Clinical Neurophysiology*, 113.844–52.

McDonald, S. A., & Shillcock, R. C. (2003). Low-level predictive inference in reading: The influence of transitional probabilities on eye movements. *Vision Research*, 43(16), 1735-1751.

Mecklinger, A., Johansson, M., Parra, M., & Hanslmayr, S. (2007). Source-retrieval requirements influence late ERP and EEG memory effects. *Brain Research*, 1172, 110-123.

Mecklinger, A., Rosburg, T., & Johansson, M. (2016). Reconstructing the past: The late posterior negativity (LPN) in episodic memory studies. *Neuroscience & Biobehavioral Reviews*, 68, 621-638.

Mirman, D., Yee, E., Blumstein, S.E., & Magnuson, J.S., (2011). Theories of spoken word recognition deficits in Aphasia: evidence from eye-tracking and computational modeling. *Brain & Language*, 117 (2), 53–68. https://doi.org/10.1016/j.bandl.2011.01.004.

Miyake, A., Carpenter, P., & Just, M. A. (1994). A capacity approach to syntactic comprehension disorders: Making normal adults perform like aphasic patients. *Cognitive Neuropsychology*, 11, 671–717.

Moreno, E.M, & Kutas, M. (2005). Processing semantic anomalies in two languages: an electrophysiological exploration in both languages of Spanish–English bilinguals. *Cognitive Brain Research*, 22(2), 205-220.

Nakano, H., Saron, C., & Swaab, T. Y. (2010). Speech and span: Working memory capacity impacts the use of animacy but not of world knowledge during spoken sentence comprehension. *Journal of Cognitive Neuroscience*, 22, 2886–2898. http://dx.doi.org/10.1162/jocn.2009.21400

Nozari, N., Mirman, D., & Thompson-Schill, S.L., (2016). The ventrolateral prefrontal cortex facilitates processing of sentential context to locate referents. *Brain & Language*, 157–158, 1–13. https://doi.org/10.1016/j.bandl.2016.04.006.

Nozari, N., Trueswell, J.C., & Thompson-Schill, S.L., (2016). The interplay of local attraction, context and domain-general cognitive control in activation and suppression of semantic distractors during sentence comprehension. *Psychological Bulletin Review*, 23 (6), 1942–1953.

Osterhout, L., & Holcomb, P. J. (1992). Event-related brain potentials elicited by syntactic anomaly. Journal of *Memory and Language*, 31(6), 785-806.

Osterhout, L., & Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language*, 34(6), 739-773.

Otten, M., & Van Berkum, J. J. A. (2008). Discourse-based lexical anticipation: prediction or priming? *Discourse Processes*, 45(6), 464–496.

Perfetti, C. A. (2007). Reading ability: lexical quality to comprehension. *Scientific Studies of Reading*, 11, 357–383. doi: 10.1080/10888430701530730

Perfetti, C. A., Wlotko, E. W., & Hart, L. A. (2005). Word learning and individual differences in word learning reflected in event-related potentials. *Journal of Experimental Psychology*, 31, 1281–1292. doi: 10.1037/0278-7393.31.6.1281

Rayner, K., & Well, A. D. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review*, 3(4), 504-509.

Rayner, K., Binder, K. S., Ashby, J., & Pollatsek, A. (2001). Eye movement control in reading: Word predictability has little influence on initial landing positions in words. *Vision Research*, 41(7), 943-954.

Regel, S., Gunter, T. C., & Friederici, A. D. (2011). Isn't it ironic? An electrophysiological exploration of figurative language processing. *Journal of Cognitive Neuroscience*, 23(2), 277-293.

Roland, D., Yun, Hongoak, Koenig, J.-P., & Mauner, G. (2012). Semantic similarity, predictability, and models of sentence processing. *Cognition*, 122, 267–279.

Rönnberg. J, Lunner. T, Zekveld. A, Sörqvist. P, Danielsson. H, Lyxell. B, Dahlström. O, Signoret. C, Stenfelt. S, Pichora-Fuller. M. K., & Rudner, M. (2013). The Ease of Language Understanding (ELU) model: theoretical, empirical, and clinical advances. *Frontier System Neuroscience*, doi: 10.3389/fnsys.2013.00031. PMID: 23874273; PMCID: PMC3710434.

Ryskin, R., Levy, R., & Fedorenko, E. (2020). Do domain-general executive resources play a role in linguistic prediction? Re-evaluation of the evidence and a path forward. *Neuropsychologia*, 136, 107258.

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. *Cognition*, 128(3), 302-319.

Spotorno, N., Cheylus, A., Van Der Henst, J. B., & Noveck, I. A. (2013). What's behind a P600? Integration operations during irony processing. *PloS ONE*, 8(6), e66839.

Staub, A. (2011). The effect of lexical predictability on distributions of eye fixation durations. *Psychonomic Bulletin & Review*, 18(2), 371-376.

Staub, A. (2015). The effect of lexical predictability on eye movements in reading: Critical review and theoretical interpretation. *Language and Linguistics Compass*, 9(8), 311-327.

Tanner, D. (2013). Individual differences and streams of processing. *Linguistic Approaches to Bilingualism* 3, 350–356.

Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 443–467.

Van Petten, C. & Luka, B. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, 83, 176–190.

Van Petten, C., Weckerly, J., McIsaac, H. K., & Kutas, M. (1997). Working memory capacity dissociates lexical and sentential context effects. *Psychological Science*, 8(3), 238-242.

Vos, S. H., Gunter, T. C., Kolk, H. H., & Mulder, G. (2001). Working memory constraints on syntactic processing: An electrophysiological investigation. *Psychophysiology*, 38, 41– 63. http://dx.doi.org/10.1111/1469-8986.3810041.

Waters, G. S., & Caplan, D. (1996). Processing resource capacity and the comprehension of garden path sentences. *Memory & Cognition*, 24, 342–355. http://dx.doi.org/10.3758/BF03213298

Yoon, J. (2008). Observations of Observations of the N400 Event Related Potential and the Gamma the Gamma Band Activities related to the familiarity during visual word perception. [Master's Thesis, Yonsei University].

Yun, H., & Hong, U. (2014). The Effect of Role Predictability and Word Predictability on Sentence Comprehension. *Journal of Cognitive Science*, 15, 349-390.

Yun, H., Lee, D., Nam, Y., & Hong, U. (2017). The predictability effect on eye movement in reading Korean dative sentences. *Language and Information*, 21(1), 73-99.