The Effects of Word Predictability and Contextual Uncertainty in the Processing of Korean Dative Sentences: An Eye-Movement Reading Comprehension Study [1]

Hongoak Yun, Dongsu Lee, Upyong Hong

Gachon University, Konkuk University, Konkuk Univeristy
oaktreepark@gmail.com, lolliriru@naver.com, uphong@konkuk.ac.kr

The purpose of this study was to investigate the conjoined roles of word predictability and contextual uncertainty in modeling readers' processing difficulty in the integration of recipients or patients into Korean dative sentences, by observing readers' eye movements in reading. The recipients and patients were arranged in a canonical order (i.e., recipient before patient) or a non-canonical order (i.e., patient before recipient). Using a cloze task, we measured a word's conditional probability to estimate the degree of predictability for a target word given a context. The outputs from the cloze-type listing task were used to compute entropy-based contextual uncertainty corresponding to the degree of contextual constraint at the point in which target words would appear. While the effect of word predictability on eye movement in reading was being taken into account in linear mixed-effect models, the significant effect of contextual uncertainty emerged only in the processing of target words in non-canonical sentences in a way that words were read more rapidly as the strength of contextual uncertainty increased. Including the contextual-uncertainty factor in the model of word predictability improved the goodness of the model fit significantly in the condition of canonical sentences and marginally significantly in that of non-canonical sentences. Our results suggest that the role of contextual

[1] This research was supported by National Research Foundation of Korea NRF-2013S1A5A2A03044144.

Corresponding authors: oaktreepark@gmaill.com (Hongoak Yun) and uphong@konkuk.ac.kr (Upyong Hong). We appreciate anonymous reviewers for their precious comments.

Journal of Cognitive Science **18-2: 127-152, 2017**
Date submitted: 05/31/17 Date reviewed: 06/07/2017
Date confirmed for publication: 06/28/2017

uncertainty is important to improve the model performance accounting for expectation-based probabilistic readers' behaviors during sentence comprehension. A weak context (i.e., high contextual uncertainty) in which a wide distribution of possible choices is constructed contributes to the reduction of processing difficulty, in particular, when readers do not make a strong commitment for upcoming structural information.

Keywords: *expectation, predictability, uncertainty, entropy, sentence comprehension, eye-movement in reading*

1. Introduction

One of the recent approaches in sentence comprehension examines readers' behaviors under the assumption that human minds (as in comprehenders' mind) are apt to compute immediately what is likely to occur next in a given context (Chater & Manning, 2006; Hale, 2001; Jaeger, 2010; Jurafsky, 1996; 2003; Levy, 2008). In this approach, important questions include which aspects of expectation need to be attentive on, how the strength of expectation is quantifiable, and how many expectation measurements are called for accounting for readers' probability-based behaviors.

One view has proposed that the degree of processing difficulty of a word in a sentence is entirely determined by the degree to which the word (or a syntactic category associated with the word) would likely to appear in a given context (Hale, 2001; Levy, 2008). For example, given the sentence fragments like (1), the processing difficulty associated with *the proposal* is determined by how likely the noun phrase would appear in the context. This claim is supported by a number of studies that have successfully demonstrated readers' cognitive effort with the conditional probability of a word (or syntactic categories of the word) being a main predictor (Ashby, Rayner, & Clifton, 2005; Bicknell, Elman, Hare, McRae, & Kutas, 2010; Boston, Hale, Patil, Kliegl, & Vasishth, 2008; Boston, Hale, Vasishth, & Kliegl, 2011; DeLong, Urbach, & Kutas, 2005; Demberg & Keller, 2008; Hale, 2001; Levy, 2008; Pado & Crocker, 2009, Roland et al., 2012; Staub, 2010; 2011).

(1) He accepted the proposal [2] …..

Another view claims that it is also important to consider how strongly readers feel it certain or uncertain to encounter what is upcoming, regardless of how strongly they expect to encounter a particular choice. For example, the processing difficulty associated with *the proposal* in Example (1) is also influenced by how uncertain a choice would be selected out of all possible choices (e.g., *the proposal*) at that position, independently of how likely *the proposal* would occur for that position. Previous findings have revealed that the effect of uncertainty is additional to that of word predictability in accounting for readers' processing difficulty (Frank, 2013; Hale, 2006; Linzen & Jaeger, 2015; Roark, Bachrach, Cardenas, & Pallier, 2009; Yun, Chen, Hunter, Whitman, & Hale, 2015).

Testing the role of uncertainty in comprehension was initiated by measuring the extent of structural uncertainty based on the probability distribution of possible structural choices. For example, according to the *entropy-reduction hypothesis* (Hale, 2006; Yun et al., 2015), the degree of processing difficulty associated with W_n (word at the n^{th} position) increase more as the degree of structural uncertainty computed at W_n is increasingly reduced in comparison to that of structural uncertainty computed at W_{n-1}(word at the $n\text{-}1^{th}$ position). According to the *competition hypothesis* (Elman, Hare, & McRae, 2005; Roark et al., 2009), the extent of processing difficulty associated with W_n (word at the n^{th} position) increase more as the degree of structural uncertainty at W_n is larger.

Should the uncertainty be the matter of possible structures? What if structural uncertainty associated with W_n is equated at W_{n-1}, as shown in the examples (2-3)? Neither the *entropy-reduction hypothesis* nor the *completion hypothesis* does predict processing differences at W_n between the sentences, as long as their claims are rooted on the uncertainty of possible structural choices corresponding to W_n. Then, what about testing the role of the uncertainty which is based on possible word choices at W_n? We think that the uncertainty based on possible word choices is likely to

[2] The example sentence was mentioned in the Linzen and Jaeger's (2015) study.

quantify the strength of contextual constraint (Yun, Mauner, Roland, & Koenig, 2012).

(2) The aboriginal man attacked the lion with a/an [W_n]
(3) The aboriginal man jabbed the lion with a/an [W_n]

In this paper, we aimed to investigate the in-depth relationship between word predictability and contextual uncertainty at W_n, especially when the degree of structural uncertainty associated with W_n is equated at W_{n-1} but the structural expectation for an upcoming position at W_n is satisfied or failed. For this purpose, we compute not only the predictability of a word for a given context but also the degree of uncertainty based on possible word choices. Our ultimate goal is to explore the underlying relationship between word predictability and contextual uncertainty in the integration of a word into a sentence as readers incrementally establish their expectation while reading dative sentences in Korean.

1.1 Literature review

Equation (1) represents the claim of the *surprisal model* such that the degree of difficulty in the integration of a word into a sentence is proportional to the degree to which the word is expected given the context. For example, in sentences like *The horse raced past the barn fell*, readers have extreme difficulty in the processing of the unexpected main verb, *fell*, into the sentence fragment. Hale (2001) succeeded in demonstrating that the degree of surprisal (i.e., negative log-transformed conditional probability) corresponding to the verb, *fell*, sharply increased, indicating that readers' difficulty could be accounted for by the fact that readers did not expect to encounter the main verb, *fell*, at the sentence-final position.

$$difficulty \propto -\log P\ (w_i\ |w_{1\ldots i-1}, \mathrm{CONTEXT}) \qquad \text{Equation (1)}$$

Equation (2), known as *Shannon's Entropy*, indicates the extent of uncertainty based on the probability distribution of possible choices that could occur at time *t* at a sentence. Several studies have investigated how much the extent of uncertainty could be reduced word by word as the

information that each word conveys is incrementally built up. This is the probabilistic measurement indicating that the difficulty in the integration of a word into a sentence is affected by the extent of uncertainty that exists at the position that the word appears.

$$H\ (X; w_1^t) = -\sum_{x \in X} P\ (x|w_1^t)\ \log \mathrm{P}(\mathrm{x}|w_1^t) \qquad \text{Equation (2)}$$

So far, two approaches have attempted to explore how the entropy-based uncertainty accounts for readers' behavior. One approach, called the *Entropy Reduction Hypothesis*, aims to test that as each word is cumulatively introduced in a sentence, the extent of uncertainty on the upcoming information is reduced which in turn would contribute to the increase of processing difficulty (Hale, 2006). Equation (3) represents the reduced amount of uncertainty at time *t*+1 in a sentence after a word that has occurred at time *t* has been processed. Existing studies demonstrated that the difficulty of processing a word in a sentence is positively correlated with a proportion to the reduced extent of uncertainty between at the very point that the word occur and at the point that the previous word has been introduced, by computing the extent of uncertainty based on part-of-speech assignments (Frank, 2010), and context-free syntactic tree structures (Hale 2006; Linzen & Jaeger, 2015; Yun et al., 2015). The *entropy reduction hypothesis* was supported by behavioral studies which observed the positive correlation between reading times and the extent of entropy computed based on individual words (Frank, 2013).

$$\Delta H\ (X; w_{t+1}) = H(X;\ w_1^t) - \mathrm{H}\ (\mathrm{X};\ w_1^{t+1}) \qquad \text{Equation (3) } 1\ ^{3}$$

Another view, called the *competition hypothesis*, is to take uncertainty as a key factor in processing difficulty in terms of competition. The hypothesis argues that having a number of possible candidates increases the degree of uncertainty and the strength of the competition, driven by high uncertainty, results in the increase of processing difficulty (Elman et al., 2005; McRae, Spivey-Knowlton, & Tanenhaus, 1998). In this approach, studies have more

[2] Equation (2) and Equation (3) are taken from Frank (2013).

focused on the amount of entropy itself rather than that of reduced entropy and showed positive correlation between the amount of entropy and reading times (Roark et al., 2009).

Unlike most studies that used entropy to compute the extent of uncertainty based on the probability distribution of possible structural choices, Yun et al. (2012) applied the entropy-based uncertainty to measure the degree of uncertainty based on possible word choices (e.g., NP as a highly likely choice in *The man jabbed the lion with a …*). They then defined the degree of uncertainty based on possible word choices to quantify the strength of a context. Yun et al. (2012) observed the additional processing facilitation due to shared semantic information among possible word choices, so-called semantic similarity, only when the contextual uncertainty was high (i.e., when the context was weakly constraining) but not when it was low (i.e., when the context was strongly constraining). The authors claimed that when readers' expectation was not satisfied, readers might feel it easier to overcome their surprise for the case in which the expectation for the upcoming information was relatively weak than for the case in which it was extremely high. That is, strong expectation for a particular choice tends to elicit processing advantages when it is satisfied but engender processing disadvantages when it is violated. In a similar vein, weak expectation can play a role of being disadvantageous since readers are unsure of the upcoming information and being advantageous since they might feel it easy to overcome even when their expectation is not met.

1.2 Studies in Korean

Unlike the head-initial languages like English in which verbs appear prior to their arguments, verb argument information does not play a significant role in cueing an upcoming argument in head-final languages like Korean and Japanese in which verbs appear sentence finally after arguments. However, the order of arguments (e.g., canonicality) often has a crucial effect on the prediction of how upcoming words associated with arguments are formulated in the downstream of the sentences. In fact, the canonicality effect of dative constructions was found in Japanese. For example, more anticipatory looks toward THEME corresponding to Direct Object emerged in the dative construction than in the accusative condition (Kamide,

Altman, & Haywood, 2003). P600 amplitudes, an ERP component referring to neural responses which are known to correspond to the difficulty in structural integration, were bigger in the accusative condition where themes appeared before recipients than in the dative condition where recipients appeared before themes (Yasunaga, Muraoka, & Sakamoto, 2010). These studies suggest that accusative-marked NP and verbs were predicted after dative-marked NPs in the dative condition, whereas only verbs were predicted after accusative-marked NPs in the accusative condition.

In Korean, the canonicality effect of dative constructions was also observed. Using the Sejong Corpus, sentences of canonical order occurred more frequently than sentences of non-canonical order (Choi, 2007). The sentences of canonical order were read faster than sentences of non-canonical order using the self-paced reading paradigm (Yun & Hong, 2014) and the eye-tracking reading paradigm (Yun, Lee, Nam, & Hong, 2017). Using the stops-making-sense reading paradigm, Yun and Hong (2014) observed the independent effect of word predictability in addition to that of role predictability, meaning that highly predictable words were read faster than less predictable words both when the roles associated with those words were highly likely (e.g., canonically-ordered sentences) and when they were unlikely (e.g., non-canonically-ordered sentences). Moreover, using an eye-movement reading paradigm, Yun et al. (2017) replicated Yun et al.'s (2014) results. They further showed that such predictive processing occurred in the early stage of sentence processing (i.e., first-pass fixation durations rather than regression fixations) and that the effect of word predictability was stronger when roles corresponding to the words were unlikely (non-canonical sentences), compared to when they were highly likely (canonical sentences).

1.3 Research questions, hypotheses, and predictions

The research questions for this study are as follows:

Question 1: Would the effect of contextual uncertainty appear independently when the effect of word predictability is taken into account? If so, how and in which way does the effect of contextual uncertainty emerge?

According to the *competition hypothesis*, it is hypothesized that the degree of processing difficulty would be positively correlated with contextual uncertainty as the degree of contextual uncertainty increases (i.e., as the context becomes more uncertainty). This is because the *competition hypothesis* claims that the strength of competition would increase as the context becomes uncertain which in turn leads to higher processing difficulty due to the increased competition.

Alternatively, according to the *context constraint hypothesis* (Yun et al., 2012), it is expected to observe that the degree of processing difficulty will be negatively correlated with the contextual uncertainty as the degree of the contextual uncertainty increases. This is potentially due to the additional facilitation of other possible words that might occur instead of the target by the degree to which the target was semantically similar to the other possible choices.

Question 2: When would the effect of the contextual uncertainty be more powerful – when structural expectation is satisfied or not?

According to Yun et al. (2017), much bigger effect of word predictability emerged on readers' first pass fixation durations on target words when recipients and patients were canonically ordered (i.e., when structural expectation was fully satisfied) but not when the arguments were non-canonically ordered (i.e., when structural expectation was dramatically failed). The authors claimed that when structural expectation (namely role predictability) was fully satisfied, it was not extremely important to encounter which role filler would occur. However, when structural expectation was failed, encountering predictable role fillers led to the significant reduction of processing difficulty in the integration of a word into a sentence.

If the reasoning claimed in the Yun et al.'s (2017) study is relevant to the current study, it is expected that the effect of contextual uncertainty would be more powerful in the non-canonically ordered sentences of low structural probability than in the canonically ordered sentences of high structural probability, although the structural uncertainty across the two

types of context was equivalent. This could mean that, when structural expectation (i.e., role predictability) is fully satisfied, it is not extremely important to encounter whether or not the probabilities of possible word choices are widely or narrowly distributed. However, when structural expectation fails, it might be relatively important to encounter the context in which the probabilities of possible word choices are somewhat narrowly distributed, independently of how predictable a particular word choice is for a given context.

2. The current study

2.1 Structural and contextual uncertainty in Korean dative sentences

Structural uncertainty. To compute the degree of structural uncertainty at the position in which words corresponding to patients or recipient roles would appear, we used the completions from the cloze task using the sentence fragments like (4) and (5). Forty undergraduate students participated in this cloze task and their completions were coded by their structural and thematic roles.

(4) *Chelwu-ka* *kyengchal-eykey* _____
Chelwu-$_{NOM}$ policeman-$_{DAT}$
Chelwu a policeman
(5) *Hochel-ika* *sinpwuncung-ul* _____
Hochel-$_{NOM}$ ID card-$_{ACC}$
Hochel (his) ID card

Using the function in Equation (4), the results of the cloze task are displayed in Figure 1a-b and Table 1. Note that the measurements were estimated under the assumption that the sentence fragments like (4) continued with themes/patients and that the sentence fragments like (5) continued with recipients.

(a) given the sentence fragment (4) (b) given the sentence fragment (5)

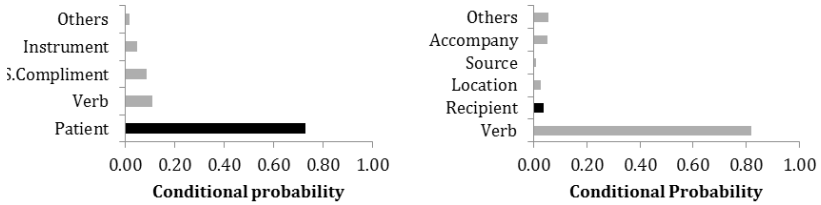


Figure 1a-b. (a) refers to the probability distribution of possible choices when an agent and a recipient were presented, whereas (b) refers to the probability distribution of possible choices for the upcoming position when an agent and a patient/theme were presented. S. Compliment refers to sentential complement. Black bars refer to target constituents, whereas gray bars refer to other possible constituents that could occur.

Table 1. The predictability-related measurements computed from the responses obtained by the cloze and listing tasks using the sentence fragments like (4-5)

	Agent + recipients _____ (e.g., sentence fragments in (4))	Agent + patients _____ (e.g., sentence fragment in (5))
Target Probability	.73	.03
Structural uncertainty at targets	.67	.63

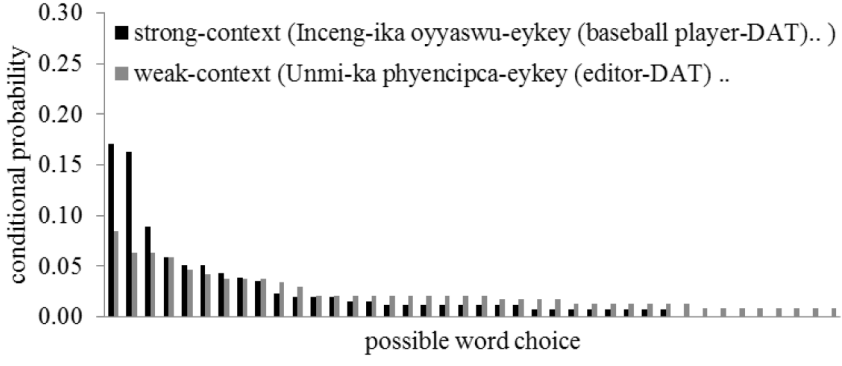
It was of great interest to find that structural probability to encounter targets was enormously different across the two conditions (i.e., .73 vs. .03). Given the sentence fragments like (4), after an agent and a recipient were presented, it was highly likely to encounter a theme/patient immediately. On the other hand, given the sentence fragments like (5) in which an agent and a theme/patient were presented, what readers were mostly likely to encounter was a verb but not a recipient. Importantly, the degree of structural uncertainty was not different between the sentence fragments at W_{n-1} (i.e., .67 vs. .63). The uncertainty at W_{n-1} in (4) was relatively low because there was high probability of encountering a NP of theme/patient,

whereas the uncertainty at W_{n-1} in (5) was relatively low because there was high predictability to encounter a verb. In sum, as for the two sentence fragments, the probability distribution of possible structural choices was equated but the probability of a target structure was enormously different.

Contextual uncertainty. To compute the degree of uncertainty based on possible words at the position in which words corresponding to patients or recipient roles was assumed to appear, we used the completions from the listing task in which participants were asked to list possible patients given the sentence fragments like (4) or to list possible recipients given the sentence fragments like (5). Forty undergraduate students participated in this listing task.

Using the function in Equation (2), the uncertainty of contextual uncertainty was computed. Figure 2a-b reveals the examples of the probability distribution of possible word choices as a function of the strength of contextual uncertainty. In the both figures, the probability distribution of possible word choices shows a clear distinction between when the context was strongly constraining like *Incheing-ika oyyaswu-eykey* (Incheing$_{NOM}$ baseball player$_{DAT}$) in Figure 2a or *Minkyu-ka welsey-lul* (Minkyu$_{NOM}$ rent$_{PAT}$) in Figure Figure 2b versus when the context was weakly constraining like *Unmi-ka phyencipca-eykey* (Unmi$_{NOM}$ editor$_{DAT}$) in Figure 2a or *Unswu-ka Sayngswu-lul*(Unswu$_{NOM}$ bottle water$_{PAT}$) in Figure 2b.

(2a) Probability distribution of possible patients given the sentence fragment (4)



(2b) Probability distribution of possible recipients given the sentence fragment (5)

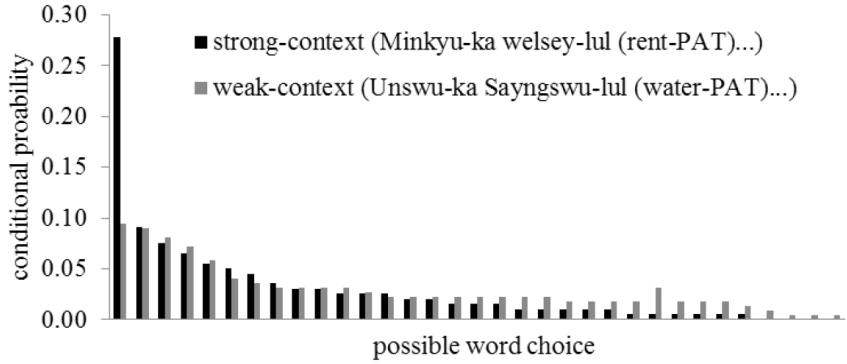


Figure 2a-b. (2a) refers to the probability distribution of possible word choices when a patient was assumed to appear after an agent and a recipient, whereas (2b) refers to the probability distribution of possible word choices when a recipient was assumed to appear after an agent and a patient. In each figure, the probability distribution of the strongly constraining context was colored in black, whereas that of the weakly constraining context was colored in gray.

As found in the Yun et al.'s (2012) study, we also observed similar characteristics that might function to hint to what degree the context was constraining. In the strongly constraining context, colored in black in Figure 2 a-b, there was a particular choice that was highly expected to appear next for a given context, relative to other possible choices. There was a sudden drop of word probabilities from highly likely choices to medium or unlikely choices. In contrast, in the weakly constraining context, colored in gray in Figure 2a-b, there was no extreme expectation for a particular choice, resulting in a gradual and wider range of probability distribution of possible choices. Regardless of whether target arguments were arranged in canonically or non-canonically, we observed the different variances of the contextual uncertainty.

2.2 Online Experiment

We used the eye-tracking reading data that Yun et al. (2017) collected. Thus, the dependent variables that we used for this study were the same with the Yun et al.'s (2017) study. Unlike the previous study, however, our new goal was to examine the effect of contextual uncertainty based on possible word choices when structural uncertainty was equated when the effect of word predictability was being taken account. At first, we summarized the online study that Yun et al. (2017) has conducted.

Participants. Seventy-four university students attended in the experiment. Each was paid at Korean 10,000 won (approximately equivalent to US $10) for their participation.

Material. Twenty-four sets of experimental stimuli were used. To prevent from readers' preview of target words (i.e., words corresponding to patients and recipients), Yun and Hong's (2014) materials were slightly modified by having adverbs between the two internal arguments. A set of example sentences were shown the sentences in (6a-d).

(6a) High Role Predictability, High Word Predictability:
Chelwu-ka | *kyengchal-eykey* | *tangtanghakey* | ***sinpwuncung-ul*** | *ceysi-hayss-ta*
Chelwu-$_{NOM}$ policeman-$_{DAT}$ proudly ID card-$_{ACC}$ showed
Chelwu showed a policeman (his) ID card proudly.

(6b) High Role Predictability, Low Word Predictability:
Minhoka | *moteyl-eykey* | *tangtanghakey* | ***sinpwuncung-ul*** | *ceysi-hayss-ta*
Minho-$_{NOM}$ model-$_{DAT}$ proudly ID card-$_{ACC}$ showed
Minho showed a model (his) ID card proudly.

(6c) Low Role Predictability, High Word Predictability:
Hochel-ika | *sinpwuncung-ul* | *tangtanghakey* | ***kyengchal-eykey*** | *ceysihayssta*
Hochel-NOM ID card-ACC proudly policeman-DAT showed
Hochel showed (his) ID card to a policeman proudly.

(6d) Low Role Predictability, Low Word Predictability:

Wuseng-ika | *sinpwuncung-ul* | *tangtanghakey* | ***moteyl-eykey*** | *ceysihayssta*
Wuseng-$_{NOM}$ ID card-$_{ACC}$ proudly model-$_{DAT}$ showed
Wuseng showed (his) ID card to a model proudly.

It was a 2 (role predictability) x 2 (word predictability) within-subject designed study. As for the role predictability factor, encountering patients/themes after agents and recipients (M = .73, $S.D.$ = .10), as in (6a-b), was more likely than encountering recipients after agents and patients/themes (M = .03, $S.D.$ = .05), as in (6c-d). The degree of word predictability associated with target words was either highly likely, as in (6a) and (6c) or unlikely, as in (6b) and (6d). The conditional probability for targets differed between predictable words (M = .15, SD = .07) and unpredictable words (M = .003, SD = .0008) in the canonical condition, that is, when roles were strongly expected (t (23) = 9.90, p < .001). Similarly, the conditional probability for target words differed between predictable (M = .17, SD = .08) and unpredictable words (M = .001, SD = .0004) in the non-canonical condition, that is, when roles were weakly expected (t (23) = 9.18, p < .001). However, the degree of word predictability was equated across the role predictability conditions (t (23) = - 1.06, p > .05). Likewise, the conditional probabilities of unpredictable words were equated, regardless of whether role predictability was high or low (t (23) = 1.30, p > .05).

The lexical properties associated with targets were also controlled to ensure that any differences of processing across the conditions could be only due to the differences of predictability but not due to the differences of other properties such as lexical frequencies, word lengths, or plausibility ratings. First, the log-transformed lexical frequencies of predictable words were not different across the conditions of role predictability (t (23) = - 0.07, p >.05). This was the same for unpredictable words (t (23) = .41, p >.05). Second, plausibility ratings associated with target words were equated when they were predictable (t (23) = .22, p >.05) and when they were unpredictable (t (23) = .57, p >.05). Third, target words were equally long between predictable and unpredictable words in the condition of low role predictability (t (23) = .37, p >.05). Words associated with recipients were

longer than words associated with patients when role predictability was high (t (23) = - 8.0, p >.001) and when it was low (t (23) = - 6.97, p >.001). We also included the length factor in our statistical models to control for any other potential effect due to the length differences across target words.

Twenty-four sets of experimental materials were counterbalanced across four presentation lists, by using a Latin-squared method and they were pseudo-randomly intermixed with 77 filler sentences. Each participant was supposed to read 101 sentences in total. The fillers had various forms of syntactic structure such as relative clauses, complex sentences, and simple declarative sentences. To keep readers' attention on reading, comprehension questions were inserted every three or four sentences. 30% of the total trials were followed by comprehension questions.

Procedure. The reading study was conducted by using Experiment Center provided by the SMI. Participants were seated in front of a 19″ display in the distance of 70cm (27.55″) between participant's eyes and the monitor display and they were asked to minimize their head movements. Eye movements of participants' left eyes were recorded by using a SMI RED 500 at the sampling rate 250Hz (viewing was binocular). All sentences were presented in a single line with a fixation marker (+) appearing at a starting point. After reading a trial, participants were instructed to fixate their eyes on the 'next' marker in the bottom of the screen to continue reading the next trial. Five practice trials were provided before the main experimental session started. The calibration of participants' eyes was checked every 10 trials.

Analysis. Four types of eye movements per word were measured: first-pass fixation duration, second pass fixation duration, total time, and regression rates. Among these, we particularly focused on the analysis of first pass fixation durations that were the sum of fixation durations on the word before leaving it for the first time. These fixation durations were known as the measurement to represent readers' early processing and their initial processing difficulty (Straub & Rayner, 2007). Besides, consistent to the other studies showing the predictability effect on the early processing measurements, Yun et al. (2017) also observed the significant effect of word predictability on first pass durations and total times, but not the other measurements. Fixation durations less than 100ms were removed

because this short time was not taken for readers' careful processing. Eye movement measurements for target words (i.e., 4th words in our materials) were submitted to a linear mixed-effect regression model for analysis. Analyses were conducted using the lme4 (version 0.999375-33, Bates & Maechler, 2014) and languageR libraries (version 1.0, Baayen, 2013) for the R statistics program (R Development Core Team, 2014).

Our analyses were conducted in two steps. First, we split the data into the group of high role predictability (canonical condition) and that of low role predictability (non-canonical condition). Then, we did a baseline analysis including factors like Word Predictability, Length, and log-transformed Lexical Frequency. Second, to test the effect of the contextual uncertainty, we added the context factor and established models of the four fixed factors: Word Predictability, Contextual uncertainty, Length, and log-transformed Lexical Frequency. Finally, we tested the model fit by comparing the model of three factors to that of four factors to check whether including the contextual uncertainty would improve the model performance in accounting for readers' behaviors.

Word Predictability, measured in cloze probability, referred to how likely target words were in their given sentence contexts. Contextual uncertainty, measured by using Shannon's entropy function, referred to how uncertain a given context would be. Although the length of target words were controlled to be equal across the conditions, Length was included to control for the effect associated with readers' perceptual efforts on length in sentence processing (Juhasz & Rayner, 2003). Finally, Frequency was the lexical frequencies of the target words. It was included in our models due to the concern that fixation durations for words were inversely proportional to their lexical frequencies, even in neutral contexts (Juhasz & Rayner, 2003; Kliegl et al., 2004; Raney & Rayner, 1995). Because any interactions among fixed factors were not of our interest, we did not include them in our models.

All fixed factors were centered. Participants and items were included as random factors together with other fixed factors. The fully-crossed and fully-specified random effect structure was simplified to yield the maximally justified random structure (Baayen, Davison, & Bates, 2008; Barr, Levy, Scheepers, & Tily, 2013). Approximately 3% of the overall

data were removed from the final model by Baayen et al.'s (2008) outlier removal procedure. Specifically, reading times with a standardized residual at a distance greater than 2.5 standard deviations from zero (in other words, means) were removed.

Results and Discussion. Recall that Yun et al. (2017) reported that the effect word predictability effect emerged only on first pass fixation durations. As expected, we also observed the effect of our testing factors on first pass fixation durations but not on the other measurements. Thus, for a simple presentation of results, we report the results of the models that contained the effect of contextual effects using first pass fixation durations: one from the canonically-ordered sentences and the other from the non-canonically ordered sentences. The results from the two models are displayed in Table 2-3, respectively, when the role predictability associated with W_n was satisfied and when it was failed.

Table 2. Results from a linear mixed-effect regression model on readers' first pass fixation durations of target words in the canonically-ordered sentences (i.e., when role predictability was high)

	Estimates	S.E.	*t*-value
Intercept	214.14	6.69	31.66
Word predictability	-21.54	38.55	-.56
Uncertainty	-9.22	8.68	-1.06
Length	5.32	11.23	.47
(Log)Frequency	2.90	8.85	.33

Note. If the absolute t-value of a fixed factor was over 2, the effect of the factor was considered to be significant at $\alpha < .05$ (Gelman & Hill, 2007).

Length and Lexical Frequency. No significant effect of Length and Lexical Frequency was found.

Word Predictability. The effect of Word Predictability was significant neither on first-pass fixation durations, meaning that a word of higher word cloze probability was not processed faster than a word of low word cloze

probability when role expectation was fully satisfied.

Contextual Uncertainty. The effect of Contextual Uncertainty was significant on first-pass fixation durations neither, meaning that the strength of contextual uncertainty, regardless of which particular word appear, did not influence readers' difficulty in the integration of a word into a sentence.

Table 3. Results from a linear mixed-effect regression model on readers' first pass fixation durations of target words in the non-canonically-ordered sentences (i.e., when role predictability was low)

	Estimates	S.E.	*t*-value
Intercept	263.97	9.92	26.62
Word predictability	-103.15	60.47	-1.71
Uncertainty	-49.78	21.12	-2.36
Length	7.97	10.89	.73
(Log)Frequency	-16.22	7.34	-2.21

Note. If the absolute t-value of a fixed factor was over 2, the effect of the factor was considered to be significant at $\alpha < .05$ (Gelman & Hill, 2007).

Length and Lexical Frequency. The effect of Length was not significant on first-pass fixation durations but that of Lexical Frequency was, indicating that lexical frequency but not length of a word had significant influences on the integration of words into sentences.

Word Predictability. The effect of Word Predictability was marginally significant on first-pass fixation durations, meaning that a word of higher word cloze probability was processed marginally faster than a word of low word cloze probability.

Contextual Uncertainty. The effect of Contextual Uncertainty became significant on first-pass fixation durations, meaning that the strength of contextual constraint, regardless of which particular word appear, did influence readers' difficulty in the integration of words into sentences. The negative sign of the coefficient indicated that the processing of recipients followed by patients became increasingly easier as the degree of contextual

uncertainty increased; in other words, as the degree of the contextual constraint became weaker.

Altogether, the results from the models in Table 2 and 3 revealed that the significant effect of contextual uncertainty was observed only when recipients and patients were non-canonically ordered, that is, when structural expectation associated with target words was fairly weak. Note that the asymmetric contribution of contextual uncertainty across the models was similar to what Yun et al. (2017) reported about the effect of word predictability. In that study, the effect size of word predictability was much bigger when role predictability associated with targets was weak rather than strong.

As done in Yun et al. (2017), we also compared the effect size of word predictability and contextual uncertainty as a function of the strength of role predictability. Figure 3 displays the absolute coefficient of the two factors when patients followed recipients in a canonical order (i.e., when role predictability was high) and when recipients followed patients in a non-canonical order (i.e., when role predictability was low). The effect size of contextual uncertainty was much bigger in the non-canonical condition (i.e., low role predictability) than in the canonical condition (i.e., high role predictability).

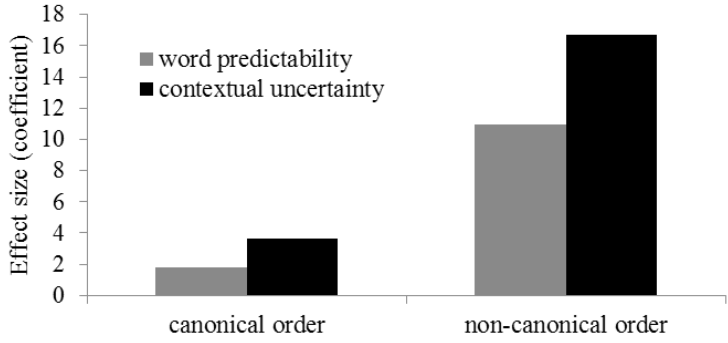


Figure 3. The effect size of word predictability and contextual uncertainty obtained from standardized models in which recipients and patients were canonically ordered and non-canonically ordered

Finally, we tested whether including the contextual uncertainty factor would attribute to improving the goodness of a model fit. For this, we compared the log-likelihood of the model which did not include the uncertainty factor to that of the model which included the context factor. First, when role predictability was high, the model fit with the contextual uncertainty factor (log-likelihood = - 4404) was significantly improved, compared to the model fit without that factor (log-likelihood = - 4405.9), χ^2 (1) = 3.91, $p < .05$. Second, when role predictability was low, the goodness of the model with the contextual uncertainty factor (log-likelihood = - 4349.3) was marginally improved, compared to that of the model without that factor (log-likelihood = - 4350.9), χ^2 (1) = 3.36, $p = .06$.

In sum, when the effect of word predictability was being taken into account, the additional effect of the contextual uncertainty was significant only when role predictability associated with target words was low. Moreover, the effect size of the contextual uncertainty was bigger for the model in which role predictability was low than for the model in which role predictability was high. Regardless of whether the additional effect of the contextual uncertainty was significant, adding the factor improved the goodness of the model fit.

3. General Discussion

The purpose of this study was to examine whether the effect of the contextual uncertainty based on possible words would be significant independently of that of word predictability, especially when the degree of structural uncertainty associated with target words was equated but that of structural predictability associated with them was different. In this section, we discuss our results by answering our questions that were raised above.

Would the effect of the contextual uncertainty appear independently when the effect of word predictability is taken into account? If so, how and in which way does the effect of contextual uncertainty emerge?

Using eye movements, we observed the additional significant effect of the contextual uncertainty only for the model in which role predictability was low (i.e., arguments were presented in a non-canonical order; e.g., patients before recipients). The effect of the contextual uncertainty indicated that the processing difficulty associated with recipients following patients was reduced as the degree of the contextual uncertainty increased, namely, as the strength of the contextual constraint was weaker.

Recall our hypotheses. According to the *competition hypothesis*, the degree of processing difficulty would be positively correlated with the contextual uncertainty as the degree of the contextual uncertainty increases. This is because the strength of competition would increase as the context becomes uncertain, which in turn leads to higher processing difficulty due to the increased competition. Alternatively, the *context constraint hypothesis* predicts that the degree of processing difficulty should be negatively correlated with the contextual uncertainty, potentially due to the additional facilitation of other possible words that might occur instead of the target by the degree to which the target was semantically similar to the other possible choices.

The negative correlation between the contextual uncertainty and the reading times that we observed in our study supported the *contextual constraint hypothesis* but not the *competition hypothesis*. Note that since we did not test the role of semantic similarity as a function of the

contextual uncertainty, the findings of the current study is not exactly same with the Yun et al.'s (2012) ones in which the additional facilitation due to shared semantic information emerged in the weakly-constraining context but not in the strongly-constraining context. Nonetheless, we agree with the underlying mechanism with what Yun et al. proposed. That is, when the context was weakly constraining (i.e., the degree of the contextual uncertainty was relatively low), readers might have constructed a wide distribution of possible choices with no strong commitment to an upcoming word. Then, with no particular expectation, the integration of a word into a sentence would be easy if the context allows a wide and less constraining distribution for possible choices, probably due to the additional facilitation from other possible choices.

When would the effect of the contextual uncertainty be more powerful – when structural expectation is satisfied or not?

The effect size of the contextual uncertainty was much bigger for the model in which role predictability was low than for the model in which role predictability was high. However, regardless of whether the contextual uncertainty was additionally significant, including the contextual uncertainty improved the goodness of the model fit.

Recall that Yun et al. (2017) showed much bigger effect of word predictability on readers' first pass fixation durations on target words when recipients and patients were canonically ordered (i.e., when structural expectation was fully satisfied) but not when the arguments were non-canonically ordered (i.e., when structural expectation was dramatically failed). They argued that, when structural expectation was fully satisfied, it was not highly important to encounter which word would occur. However, when structural expectation failed, encountering a predictable word led to the significant reduction of difficulty in the integration of a word into a sentence.

We think that the reasoning that was claimed in the Yun et al.'s (2012) study is relevant in the current study. When structural expectation (namely role predictability) is fully satisfied, it would not be extremely important to encounter whether or not the probabilities of possible word choices

are widely or narrowly distributed. However, when structural expectation failed, it might be relatively important to encounter the context in which the probabilities of possible word choices are somewhat narrowly distributed, independently of how predictable a particular word choice would be for a given context.

4. Conclusion

The purpose of this study was to investigate the conjoined roles of word predictability and contextual uncertainty during sentence comprehension. By modeling readers' processing difficulty in the integration of recipients or patients into Korean dative sentences, we found that the both types of probabilities measures contributed to accounting for readers' eye movements in reading. However, while the effect of word predictability on eye movement in reading was being taken into account, the significant effect of contextual uncertainty emerged only in the processing of targets in the non-canonical sentences in a way that words were read more rapidly as the strength of contextual uncertainty increased. Taken together, our results suggest that the role of contextual uncertainty is important to improve the model fit that account for expectation-based probabilistic readers' behaviors during sentence comprehension. The context in which allows a wide distribution of possible choices helps reduce processing difficulty, in particular, when readers do not make a strong commitment for the upcoming information. The underlying mechanism about whether or not the processing facilitation observed in the weakly constraining context might be due to the additional activation from other choices that share semantic information calls for a further study.

References

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

Baayen, R. H. 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press.

Baayen, R. H. 2013. languageR: Data sets and functions with "Analyzing Linguistic

Data: A practical introduction to statistics". R package version 1.4.1. http://CRAN.R-project.org/package=languageR

Baayen, R. H., Davidson, D. J., & Bates, D. M. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language*, *59*(4), 390-412.

Barr D. J., Levy R., Scheepers C., & Tily H. J. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255–278 10.1016/j.jml.2012.11.001.

Bates, D., & Maechler, M. 2010. lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-33. Retrieved July 9, 2014.

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, 489–505.

Boston, M. F., Hale, J. T., Vasishth, S., & Kliegl, R. 2011. Parallel processing and sentence comprehension difficulty. *Language and Cognitive Processes*, *26*(3), 301-349.

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).

Chater, N., & Manning, C. D. 2006. Probabilistic models of language processing and acquisition. *Trends in cognitive sciences*, *10*(7), 335-344.

Choi, H. W. 2007. Length and order: A corpus study of Korean dative-accusative construction. *Korean Discourse and Cognition*, *14*(3), 207-227.

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.

Elman, J. L., Hare, M., & McRae, K. 2005. Cues, constraints, and competition in sentence processing. In M. Tomasello, & D. Slobin (Eds.), *Beyond nature-nurture: Essays in honor of Elizabeth Bates*, 111-138. Mahwah, NJ: Erlbaum.

Frank, S. L. 2010. Uncertainty reduction as a measure of cognitive processing effort. In Proceedings of t*he 2010 workshop on cognitive modeling and computational linguistics* (pp. 81–89). Uppsala, Sweden: Association for Computational Linguistics.

Frank, S. L. 2013. Uncertainty reduction as a measure of cognitive processing load in sentence comprehension. *Topics in Cognitive Science*, 5, 475–494.

Gelman, A., & Hill, J. 2007. Missing-data imputation. *Behavior Research Methods,* 43(2), 310-30.

Hale, J. 2001. A probabilistic early parser as a psycholinguistic model. In: North

American Chapter of the Association for Computational Linguistics (NAACL), Vol. 2 (pp. 1–8). Association for Computational Linguistics.

Hale, J. 2006. Uncertainty about the rest of the sentence. *Cognitive Science*, *30*(4), 643-672.

Jaeger, T. F. 2010. Redundancy and reduction: Speakers manage syntactic information density. *Cognitive psychology*, *61*(1), 23-62.

Juhasz, B. J., & Rayner, K. 2003. Investigating the Effects of a Set of Intercorrelated Variables on Eye Fixation Durations in Reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* 29(6), 1312–1318.

Jurafsky, D. 1996. A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive science*, *20*(2), 137-194.

Jurafsky, D. 2003. Probabilistic modeling in psycholinguistics: Linguistic comprehension and production. *Probabilistic linguistics*, *21*.

Kamide, Y., Altmann, G. T., & Haywood, S. L. 2003. The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and language*, *49*(1), 133-156.

Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. 2004. Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology,* 16. 262–284.

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

Linzen, T., & Jaeger, T. F. 2015. Uncertainty and expectation in sentence processing: evidence from subcategorization distributions. *Cognitive science.*

McRae, K., Spivey-Knowlton, M. J., & Tanenhaus, M. K. 1998. Modeling the influence of thematic fit (and other constraints) in on-line sentence comprehension. *Journal of Memory and Language*, *38*(3), 283-312.

Padó, U., Crocker, M. W., & Keller, F. 2009. A probabilistic model of semantic plausibility in sentence processing. *Cognitive Science*, *33*(5), 794-838.

R Core Team 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Raney, G. E., & Rayner, K. 1995. Word frequency effects and eye movements during two readings of a text. *Canadian Journal of Experimental Psychology/ Revue canadienne de psychologie expérimentale,* 49(2), 151–173.

Roark, B., Bachrach, A., Cardenas, C., & Pallier, C. 2009, August. Deriving lexical and syntactic expectation-based measures for psycholinguistic modeling via incremental top-down parsing. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1* (pp. 324-333). Association for Computational Linguistics.

Roland, D., Yun, H., Koenig, J.-P., & Mauner, G. 2012. Semantic similarity,

predictability, and models of sentence processing. *Cognition,* 122, 267–279.

Staub, A. 2010. Eye movements and processing difficulty in object relative clauses. *Cognition*, *116*(1), 71-86.

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

Staub, A., & Rayner, K. 2007. Eye movements and on-line comprehension processes. In: G. Gaskell (Ed.), *The Oxford Handbook of Psycholinguistics* (pp. 327-342). Oxford, UK: Oxford University Press.

Yasunaga, D., Muraoka, S., & Sakamoto, T. 2010. Case marker effects of Japanese in predicting following element. Bulletin in Japanese Cognitive Science Society 17, 663–669.

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.

Yun, H., Mauner, G., Roland, D., & Koenig, J.-P., 2012. The Effect of Semantic Similarity is a Function of Contextual Constraint. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (p.1191-1196). Austin, TX: Cognitive Science Society.

Yun, J., Chen, Z., Hunter, T., Whitman, J., & Hale, J. 2015. Uncertainty in processing relative clauses across East Asian languages. *Journal of East Asian Linguistics*, *24*(2).