Linguistic Research 37(Special Edition): 189-218
DOI: 10.17250/khisli.37..202009.008

L2ers' predictions of syntactic structure and reaction times during sentence processing*

Euhee Kim** · Myung-Kwan Park · Hye-Jin Seo***
(Shinhan University · Dongguk University)

Kim, Euhee, Myung-Kwan Park, and Hye-Jin Seo. 2020. L2ers' predictions of syntactic structure and reaction times during sentence processing. *Linguistic Research* 37(Special Edition): 189-218. This paper investigates how Korean L2 learners of English predict upcoming syntactic structure based on a newly received word during sentence processing. Studies like Linzen and Jaeger (2016) suggest that readers use their probabilistic inference developed by their experience of the language to which they have been exposed to predict the most appropriate syntactic structure. This study replicates the experiment for L2ers following Linzen and Jaeger (ibid.), which investigates the way of predicting syntactic structure by using the subcategorization frame of a verb to understand L1 language processing. We employ the information-complexity metrics such as surprisal, entropy, and entropy reduction to quantify the uncertainty/unexpectedness of a given word that reflects the processing difficulty during sentence processing. The results show that L2ers' tendency to read different regions of a sentence varies. Reading times are longer in the verb and the ambiguous regions of the structurally ambiguous than of the structurally unambiguous sentences. Likewise, reading times are longer in the disambiguating region of the unambiguous than of the ambiguous sentences. Reading times are also longer when the surprisal increases in the disambiguating region. Overall, the findings reveal that such information-complexity metrics as entropy reduction and surprisal play an instrumental role in accounting for the aspects of sentence processing by Korean L2 learners of English. **(Shinhan University · Dongguk University)**

Keywords sentence processing, syntactic structure, prediction, uncertainty, unexpectedness, entropy, entropy reduction, surprisal, reading times

---

* We would like to thank the anonymous reviewers of this journal for their valuable comments and feedback. Usual disclaimers apply here, of course. This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF 2017S1A5A2A01026286).

** First author

*** Corresponding authors

1. Introduction

The predictability of the next word during sentence processing is known to be a core explanatory concept of information processing (Clark 2013). Under the cumulatively predictive language processing framework, the brain makes use of the relevant contextual information in a sentence to pre-activate the possible upcoming sequence of words before they are finally perceived. The predictability in question helps people to facilitate sentence processing when their prediction actually turns out to match the perception (Lowder, Choi, Ferreira, and Henderson 2017). One relevant issue here is how readers infer ensuing syntactic structure in sentence processing. Researchers have provided strong evidence that people engage in incremental processing on the basis of each newly received word during parsing (Altmann and Kamide 1999; Federmeier 2007; Hale 2001; Levy 2008; Linzen and Jaeger 2016). Readers constantly predict upcoming syntactic structure and decide on the most likely syntactic structure based on each of the incoming words. One way of predicting appropriate syntactic structure is probabilistic inference. Readers build their probabilistic inference, which is attributed to their experiences in lifelong exposure to the language they use. For instance, readers are most likely to postulate the syntactic structure that they have been exposed to in everyday use of language. Thus, there is a causal link between predictability and processing difficulty (Boston, Hale, Kliegl, Patil, and Vasishth 2008; Demberg and Keller 2008; Jennings, Randall, and Tyler 1997; McDonald and Shillcock 2003a, b). From probabilistic inference, readers choose promising syntactic structure in proportion to its probability. This leads readers to predict more frequently experienced syntactic structure faster than less frequently experienced syntactic structure (DeLong, Urbach, and Kutas 2005; Smith and Levy 2013; Ehrlich and Rayner 1981). Word/structure predictability effects may indeed serve as the evidence to understand the nature of human cognition: how far humans predict upcoming linguistic material and how they react if newly introduced elements differ from their predictions (Bar 2007).

The paper is structured as follows. The next section reviews the previous studies of the relationship between syntactic structure and the psycholinguistic measure of reaction time in sentence processing and motivates the research agenda to be explored in this paper. Section 3 and 4 each outlines the experiments performed and reports the results of the experiments. Section 5 elaborates on the results and wraps up with a conclusion.

2. Previous studies

In line with the previous works on predictive sentence processing (Henderson, Choi, Luke and Schimidet 2017; Bell, Brenier, Gregory, Girand and Jurafsky 2009; Linzen and Jager 2016), Linzen and Jager (2016) take up subcategorization frame(s), which are the option(s) that a verb takes for their complements, to understand human language processing. For example, the verb *accept* can occur with two kinds of complements, a noun phrase (e.g., *accept a gift*) or a sentential complement (e.g., *accept that you have lost*). In reality, the verb *accept* frequently favors taking as its complement the noun phrase (NP) than the sentential complement (SC). If readers parse a partial sentence like *He accepted the proposal*, they tend to take the NP *the proposal* as the complement of *accept* according to their experiences. Therefore, when the partial sentence is followed by the word sequence *was wrong*, they should undo their initial analysis of the sentence structure. The given word sequence *was wrong* is critical to conclude that the current verb takes not the NP but the SC complement. Under the predictive parsing hypothesis, the reading times (RTs) for the disambiguating region like *was wrong* vary depending on the subcategorization frame of a verb. The verb *accept* comes with NP complement more frequently than SC complement, but the verb *prove* comes with SC complement more frequently than NP complement. RTs are likely to be longer on sentences with the verb *accept* than the verb *prove* when SC complement is taken (Garnsey, Pearlmutter, Myers, and Lotocky 1997; Trueswell, Tanenhaus, and Kello 1993). In a parallel fashion to the verb *accept*, the verb *forget* favors not SC but NP complement. However, besides NP and SC complements, *forget* also takes other kinds of complements such as a prepositional phrase (PP) and an infinitive. It means that *forget* has more options in complement-taking than *accept*. This means that the degree of uncertainty in predicting upcoming syntactic structure is greater for *forget* than *accept*. RTs are likely to be longer on sentences with *forget* than *accept* during sentence processing. In a nutshell, human behaviors such as RTs are purportedly affected by the degree of uncertainty attributed to subcategorization frame options of a verb in question. Now the question is how we can quantify the effects that such verbal syntactic features as subcategorization frame exert on sentence processing?

More recently, some researchers suggest information complexity metrics to quantify the predictability for each word of a sentence. This approach is based on information theory (Shannon 1948) and developed by combining it with a state-of-the-art method in computational language modeling. It has an advantage of generating estimates of information complexity metrics for word-by-word sentence processing. Thus, it works as an online sentence processing measure. The first metric that lately began to receive attention in theories of language processing is Entropy. Entropy is related to the degree of the uncertainty about which context sources like words and structures will be being communicated to unfold the remaining part of a sentence. Therefore, entropy is a formula for quantifying the uncertainty about a probability from the various possible outcomes of a sentence by using the Shannon entropy of the probability distribution, p:

$$H = -\sum_{i=1}^{n} p_i \log_2 p_i .$$

High entropy means that there are more potential options when predicting the remaining context sources. Entropy is maximal when potential options are uniformly distributed. On the other hand, entropy is zero when there is only one potential option. Some studies suggest that the higher entropy is, the longer RTs are (Elman, Hare, and McRae 2004; McRae, Spivey-Knowlton, and Tanenhaus 1998). This hypothesis is called the Competition Hypothesis.

The second one is Entropy Reduction, which is associated with the fluctuation of entropy on encountering each new word in sentence processing. If entropy at word w_i is defined as H_i, then entropy reduction at w_i is calculated from $H_i - H_{i-1}$. It reflects the idea that communicative uncertainty decreases when entropy decreases from one word to the next. In other words, it tends to reduce uncertainty about sentence structure since incoming words are useful to predict the upcoming sentence structure. Thus, the degree of reduction in uncertainty affects language processing. The higher entropy reduction is, the longer RTs are (Hale 2003, 2006, 2011; Yun, Chen, Hunter, Whitman, and Hale 2015). This hypothesis is known as the Entropy Reduction Hypothesis. Under the entropy reduction hypothesis, an increase of entropy from one word to the next does not affect language processing because this hypothesis claims that an increase of entropy means that even though the additional word is given, readers do not predict which syntactic structure is imminent.

Third, the more traditional measure is surprisal, which quantifies relative unexpectedness at every word in a sentence context. The recent computational language models investigate word-by-word predictability to calculate surprisal values regardless of input sentences. Surprisal is defined as the negative log probability of a word in a given context, as follows:

$$\text{Surprisal}(W_i) = -\log_2 P(W_i|W_1 \ldots W_{i-1})$$

Higher surprisal is associated with more unexpectedness about the value of W_i in a current context. Using this metric, studies have proved the correlation between surprisal and online sentence processing. Higher surprisal values have a tendency to increase RTs (Boston et al. 2008; Demberg and Keller 2008; Smith and Levy 2013) and N400 amplitudes (Frank, Otten, Galli, and Vigliocco 2015). This hypothesis is termed the Surprisal Hypothesis (Hale 2001; Levy 2005, 2008).

In brief, there are information complexity metrics to track word-by-word sentence processing that assumedly simulates the language processing by readers. Entropy and entropy reduction are related to uncertainty to predict the upcoming part of a sentence. Surprisal is related to the unexpectedness of newly encountered words during parsing.

A number of recent studies have begun to investigate whether the effect of information complexity is related to reading times (RTs) (Frank 2013; Hale 2003, 2006; Roark, Bachrach, Cardenas, and Pallier 2009; Wu, Bachrach, Cardenas, and Schuler 2010; Yun et al. 2015) and neural measures (Frank, Otten, Galli, and Villioco 2015; Willems, Frank, Nijhof, Hagoort, and van den Bosch 2016). Some studies have demonstrated that the entropy reduction hypothesis is superior in predicting the well-known phenomenon of garden path in sentence processing

(Hale 2003) and the asymmetry between subject and object relative clause (Yun et al. 2015). However, these studies just compared the rough patterns from the findings in the previous studies in light of entropy reduction to show the effect of entropy reduction during sentence processing. Other studies utilized empirical RTs to examine the relation between the information complexity metrics (surprisal, entropy, and entropy reduction) and sentence processing. Roark et al. (2009) detected a positive correlation between entropy and sentence processing, supporting the competition hypothesis. Wu et al. (2010) and Frank (2013) found a positive correlation between entropy reduction and sentence processing, supporting the entropy reduction hypothesis. However, these studies just used one of the information complexity metrics to observe the nature of sentence processing. Thus, most of the previous studies have two kinds of limits: one is that language processing was explained by the information complexity metrics without actual experiments. The other is that the competition hypothesis and the entropy reduction hypothesis have not been directly compared and studied under the same experimental environment.

To resolve these limits, Linzen and Jaeger (2016) conducted the experiment to make a comparison between the competition and the entropy reduction hypotheses with the same stimuli. Specifically, Linzen and Jaeger's study examined how readers are affected by the uncertainty about upcoming syntactic structure during sentence processing by testing the competition and entropy reduction hypotheses. They first conducted a self-paced reading experiment with English native speakers to assess the effects of entropy and entropy reduction on human RTs. There are 32 sentence pairs for the experimental materials. In each pair, one type of sentence that involved the complementizer *that* after a main verb, was called Unambiguous sentence, and the other that did not, Ambiguous sentence (64 sentences in total):

(1) Ambiguous sentence: The men discovered the island had been invaded by the enemy.
Unambiguous sentence: The men discovered <u>that</u> the island had been invaded by the enemy.

There are five critical regions for target sentences: subject, verb, ambiguous, disambiguating, and rest region. First, the subject region refers to the time that participants spent at the subject position of a sentence such as *the men* in (1). Second, the verb region points to the main verb of a sentence such as *discovered*. As uncertainty about the upcoming part of the sentence is largely engendered by the main verb, this region is theoretically the important point of interest. Third, the ambiguous region contains the subject of the clausal complement (CC), *the island*; in (1), the subject can be mistakenly taken as the object of the matrix object, or correctly as the subject of the embedded complement clause; in (1), however, the former analysis is excluded. In other words, this region is the point in a sentence where readers initially have thought it as an object NP in the ambiguous sentence of (1). Fourth, the disambiguating region involves either the two auxiliary verb sequence (had been) or the auxiliary verb and negation sequence (had not) plus the past participle form of verb (invaded). This region is also meaningful because at this point readers finally take the NP *the island* not as an object, but an embedded subject of the CC. Therefore, they reanalyze the most recent sentence structure at this point of sentence processing. The rest region includes the remaining part of sentence immediately following the disambiguating region.

Linzen and Jaeger (2016) took two kinds of approach in measuring uncertainty in experimental sentences: (i) uncertainty in single-step prediction and (ii) uncertainty in full prediction. In the case of the uncertainty in single-step prediction, the information complexity metrics were only determined by the distinctive syntactic features of main verbs in each of the sentence conditions. Human RTs for each region were estimated by the information complexity metrics of main verbs in these sentence conditions. On the other hand, in the case of the uncertainty in full prediction, the information complexity metrics in each word of a sentence were estimated from a probabilistic context-free grammar (PCFG) based on the Penn Treebank. All the sentence elements except for the main verb were unlexicalized to focus on the probability of syntactic structures rather than the effect of individual content words. For example, sentences like (1) were transformed into *DT NNS discovered (that) DT NN VBD[had] VBN[been] VBN IN DT NN*. It helps to keep the grammar rationally small because it is not easy to estimate entropy for each word in sentences using larger grammars (Roark et al. 2009). Under the uncertainty in full prediction, the value of entropy will reflect the uncertainty about the next syntactic node as well as the internal structure of that node. Hence it will track online sentence processing in human cognition. In Linzen and Jaeger's (2016) study, the results from the single-step prediction supported neither the competition or the entropy reduction hypotheses because of the lack of entropy effects. However, the results from the full prediction showed that the entropy reduction was correlated with increased RTs both in the verb region (β=0.014, p=.047) and in the ambiguous region (β =0.042, p<.001), and that the entropy was correlated with decreased RTs in the ambiguous region (β=-0.043, p<.001). These results render compelling support for the entropy reduction hypothesis rather than the competition hypothesis.

3. The current study

This study replicated the earlier study by Linzen and Jaeger (2016), which demonstrates that the degree of uncertainty about the full prediction has an impact on processing difficulty during sentence processing in English native speakers. Departing from this seminal work, however, the present study investigated the same issue for not L1 but L2 speakers, to ascertain whether Korean L2 learners of English continuously infer upcoming syntactic structure, and likewise are influenced by the probability distribution of their inferences in proportion to uncertainty about upcoming syntactic structure.

As pointed out above, researchers tried to demonstrate which measure among surprisal, entropy and entropy reduction, is more likely to reflect the behavior of human language processing (Levy and Gibson 2013; Linzen and Jaeger 2016). This is still controversial and much more studies need to be done. In addition, most previous studies focused on English native speakers rather than L2 leaners of English. This study, to the best of our knowledge, is the first attempt to assess the effects of uncertainty on RTs in Korean L2 learners of English. Linzen and Jaeger (2016) utilized a probabilistic context-free grammar (PCFG) derived from the Penn Treebank (Marcus, Marcinkiewicz, and Santorini 1993) to analyze specific classes of sentences. The PCFG would vary according to which corpus an experimenter uses. The importance of a corpus cannot be stressed too highly when we plan to examine the nature of language processing by adopting a computational analysis. Researchers often employ the Penn

Treebank because it is a large-sized corpus and is made up of wide-ranging genres, like Wall Street Journal articles, IMB computer manuals, and nursing notes, etc. These kinds of genres are familiar to English native speakers. However, we cannot be sure whether it is also suitable to investigate the behaviors of Korean L2 learners of English in sentence processing. There is indeed little chance for Korean English learners to access the language materials in the Penn Treebank corpus. This study began with the inquiry of whether the PCFG from the language materials that Korean L2 learners of English are familiar with can provide a better account for their behavior in sentence processing. In keeping with the global trend, Korean students start to learn English in elementary school and study English in class using English textbooks (Yuasa 2010). The English textbooks in Korea are thus systematically produced at different stages of English learning according to the level of students' proficiency from elementary to high school. Thus, we collected textbooks published in public education of Korea from 2001 to 2009. It is expected that the PCFG from the English textbooks published in Korea rather than the Penn Treebank corpus can better explain the behavioral responses of Korean English learners in sentence processing because it consists of English expressions that Korean students have actually experienced in school. The dataset of English textbook corpora based on the English 11 middle-school and 12 highschool textbooks published in 2001 and the English 19 middle-school and 12 highschool textbooks published in 2009 together with 2016-2018 EBS-CSAT prep English books contains 231,004 sentences (a total of 2,271,692 word tokens, and 32,270 unique words; the preprocessing (sentence segmentation, data transformation, arranging syntactic grammar rule, etc.) done for the raw text data using Python's Natural Language Toolkit (NLTK) together with the re module).

3.1 Method

3.1.1 Participants

34 undergraduates from Dongguk University in Seoul, Korea participated in this study (14 males and 20 females, mean age = 26.27, range = 20~30). They were all Korean learners of English and received KRW 10,000 per hour in reward for their participation. The experiment took almost 40 minutes per

participant to complete. Their English proficiency was also measured based on the result of a cloze test adapted from O'Neil, Cornelius, and Washburn (1981). The mean score of the cloze test was 77.5% (range = 45~95%, *SD* = 9.63). Except for one participant, the score of the cloze test was greater than 60% (mean = 79.39%, range = 60~95%, *SD* = 7.73). They could be regarded as belonging to the intermediate or higher level of proficiency (Chae and Shin 2015).

3.1.2 Stimuli

The experimental stimuli consisted of 30 sentence pairs adapted from Linzen and Jaeger (2016). Each pair involved one type of sentences with the complementizer added after a main verb, *The men discovered that the island had been invaded by the enemy*, and the other without it, *The men discovered the island had been invaded by the enemy*. Even though there were 32 sentence pairs with 32 verbs in Linzen and Jaeger (2016), we removed two verbs (*advocate* and *guarantee*) to analyze the data (60 sentences in total; Appendix A for all the stimuli). The information complexity metrics of these two verbs could not be calculated because they were absent from the vocabulary in the textbook dataset. 32 filler sentences (those sentence items that involve syntactic structure different from target sentences, such as *if*-clause, *although*-clause, coordinate clause, etc.) were also used. These experimental and filler items were pseudo-randomized to make certain that the target stimuli were always preceded by filler sentences. All the stimuli were followed by the comprehension question (i.e., *Had the island been invaded?/Did the men discover the island?*). Half of the comprehension questions were syntactically and semantically congruent with the preceding sentences and the other half were not. The target sentences were put into two lists for presentation. In each list, half of the target sentences had the complementizer and the other half did not.

3.1.3 Procedure

At the beginning of the experiment, the participants signed the consent form and then were seated in front of a personal computer in our lab. Every participant was randomly assigned to an experiment list. The experimenters guided them to read the sentences as quickly and as precisely as possible to observe their sentence processing pattern in reading. The participants did not recognize the research design until the end of the study. All the experiments were run on Ibex Farm, which is the web-based experimental presentation platform (Drummond 2013). Each sentence was presented in a word-by-word, self-paced moving window method and was followed by a comprehension question to ensure that the participants were attending to the meanings of the sentences. The answer for comprehension question was made using either an F key or a J key on the keyboard. The F key means that a comprehension question matches with the trial sentence, but the J key means that a comprehension question mismatches with the trial sentence. They did not receive any feedback on their answers. After the self-paced reading experiment, they completed the cloze test to determine their English proficiency level.

3.1.4 Data analysis

Seven participants were excluded from the statistical analyses due to their low accuracy of comprehension question responses (below 70%). Prior to the data analysis, there was data pre-processing. If raw RTs per word were less than 200 ms or more than 3,000 ms, individual words were excluded from the data analysis. RTs might vary in respect of region lengths as well as individuals' reading rates. To normalize the variations in RTs, a regression equation per participant based on region length was calculated, using all the experimental items and fillers (Ferreira and Clifton 1986; Trueswell, Tanenhaus, and Garnsey 1994). For the statistical modeling, all the raw RTs were log-transformed to decrease skewness (Baayen and Milin 2010). It was removed, so that a word's log-transformed RT might be outside 3 standard deviations (SDs) above or below a participant's grand mean. Ultimately, a residual reading time, called length-corrected log RTs (log RRTs), was obtained after subtracting RTs predicted by the regression equation for each participant from the log-transformed value (Appendices B and C for the tables of length-corrected log RTs). After the overall data pre-processing, it was 3.33% less than the experimental data.

This study analyzed three regions; the (main) verb, the ambiguous, the disambiguating regions. Although the sentences were presented visually one word at a time, each of the three regions for analysis contained several words. A word's log-transformed RT was summed across words for each region. Only the log RRTs instead of raw RTs were used to analyze the data in this study. Those RRTs for each region were analyzed through linear mixed effects models using the lme4 (Bates, Maechler, Bolker, and Walker 2014) and lmerTest (Kuznetsova, Brockhoff, and Christensen 2017) packages in the R 3.5.2 environment (R Core Team 2018).[1] The main predictors in the models were Sentence Condition (categorical predictor: ambiguous vs. unambiguous), Surprisal (numerical predictor: from 0.434 to 23.191), Entropy (numerical predictor: from 11.793 to 109.898), and Entropy Reduction (numerical predictor: from 0 to 17.091). The additional predictors in the models were Participant's English Proficiency (numerical predictor: from 1 to 34) and Stimulus Order (numerical predictor: from 1 to 64).[2] To minimize removable collinearity, the categorial predictor like Sentence Condition was transformed by effect coding (e.g. ambiguous = -.5, unambiguous = .5), and then was centered. Indeed, all the numerical predictors were mean-centered. We used a maximal model including by-participants, by-items random intercepts, by-participants random slopes for all of the predictors, by-items random slopes for the critical predictor (i.e., Sentence Condition), and their interactions. When the model fitting approach did not converge statistically, the model was gradually reduced based on the approach, i.e., the back-off procedure for the random effects structure, proposed in Bates et al. (2015). After removing the random slopes step-by-step from the highest order interactions, the model was refitted. Since entropy and entropy reduction are closely related to each other, a separate model was built to assess their effect.

1 At first, we checked the normal distribution of the raw RT and the logRRT right after transforming the raw data into the logRRT by using R functions such as qqmath and densityplot. When we selected the final linear-mixed effect model, we followed the assumptions of the linear-mixed effect model made in Gries (2009) and Winter (2019). There are roughly four assumptions made there; linearity, homoscedasticity, normal distribution, multicollinearity of residuals. With the assumptions at hand, we can check the distribution of residuals, the normal distributions of residuals, the correlations among dependent variables. For linearity and homoscedasticity, the plot(final_model) function was used. For the normal distribution of residuals, the resid(final_model); qqnorm(x); qqline(x, col=2) function was used. For multicollinearity, the sqrt(vif(model)) function was used (cf. Shin, 2019).

2 Thus, both Participant's English proficiency and Stimulus Order were included as independent variables. We in turn examined their effects on RTs.

3.2 The behavioral difference between English native speakers and Korean L2 learners of English

Before analyzing the data, we briefly compared the raw RTs of English native speakers in Linzen and Jaeger (2016)[3] and those of Korean L2 learners of English in this study to look over the difference between them. Korean L2 learners of English took longer time to read each word than English native speakers, as shown in Figure 1:

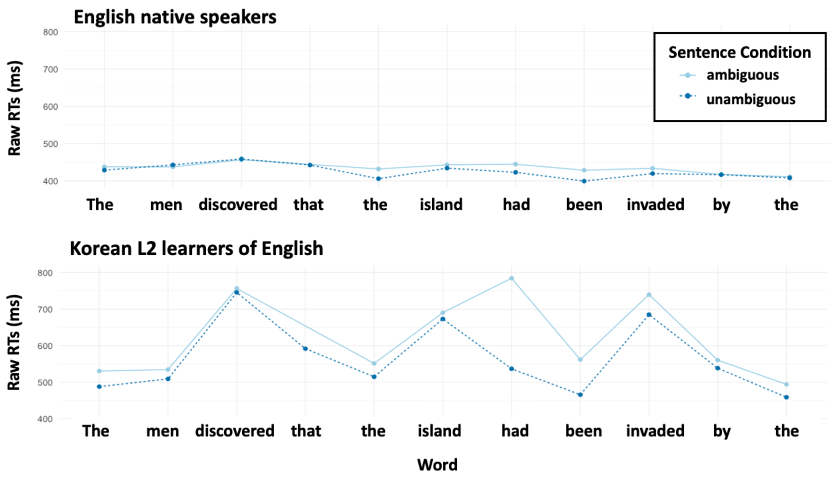

Figure 1. The mean raw reading times (RTs) in word-by-word for all the words except for the last word of sentences

There was an outstanding difference in raw RTs between them at the disambiguating region, *had been invaded*, in the ambiguous sentence condition. English native speakers seem not to have been confused when they met the unexpected complex structure. As for them, the raw RTs did not increase greatly in this region. By contrast, Korean L2 learners of English seem to have been confused in that their raw RTs increased greatly. This reading tendency shows that Korean L2 learners of English needed some time to reanalyze the sentence structure when the current structure differed from their expectation.

3 Thanks go to Tal Linzen (perl. comm.) for sharing the raw data from Linzen and Jaeger (2016).

3.3 Expected predictions for each region under each hypothesis

In the main verb region, the higher full entropy in this region is derived from a main verb with higher subcategorization entropy. In addition, the full entropy is determined by the degree of uncertainty about the internal structure after a main verb or the internal entropy estimated from the probability of the internal structure of a verb complement. As pointed out above, suppose that for example, there are two verbs, verb A and verb B, and these verbs can take SC and NP as their complements. Verb A takes SC and NP at a ratio of 7:3, while verb B does so at a ratio of 3:7. Because SC has more complex internal structure than NP or PP complement, it follows that verb A is expected to have greater/higher internal entropy than verb B. A greater internal entropy results in a greater full entropy. Consequently, verb A is greater in full entropy as well as internal entropy than verb B. Since entropy is similar across items in the subject region, entropy and entropy reduction are mutually inversed in this region. If the degree of uncertainty of a verb increases from the subject to the verb region (the verb has high internal entropy), the full entropy will be high, but the full entropy reduction will be zero. On the other hand, if the degree of uncertainty of a verb decreases (the verb has low internal entropy), the full entropy will be low but the full entropy reduction will be high. The competition hypothesis expects that in the main verb region there is a positive correlation between entropy and RTs, but the entropy reduction hypothesis predicts that in the main verb region there is a positive correlation between entropy reduction and RTs.

In the ambiguous region, the sentence conditions differ markedly in the full entropy. In the unambiguous sentences, the participants can evidently recognize the sentence structure after the complementizer *that*. It means that the probability of SC becomes 1, causing the full entropy to increase sharply. On the other hand, in the ambiguous sentences (without the complementizer *that*), the participants temporarily assume the NP complement in this region. Among several options as the complement of a main verb, the NP complement is relatively low in internal entropy, causing the full entropy to remain low. Under the competition hypothesis, RTs should be greater in the ambiguous region of the unambiguous than of the ambiguous sentences. Contrary to the competition hypothesis, under the entropy reduction hypothesis, RTs should be higher in the ambiguous region of the ambiguous than of the unambiguous sentences.

In the disambiguating region, participants get convinced that the complement structure is SC. It is expected to induce the overall high full entropy regardless of sentence condition. In the ambiguous sentences, the entropy is high in the disambiguating region due to the higher internal entropy of SC. Its degree depends on the probability of the internal structure of a verb complement. In the disambiguating region of the unambiguous sentences, the entropy will remain high but go down to a certain degree because the participants have already recognized the complement structure in the ambiguous region. Thus, under the competition hypothesis, it is predicted that there should be no big difference across sentence conditions in this region. Under the entropy reduction hypothesis, on the other hand, it is predicted that RTs should be higher in the unambiguous than in the ambiguous sentences. Recall that the entropy reduction hypothesis suggests that there is no processing cost incurred when the entropy increases. Hence, no processing difficulties are expected to arise in this region of the ambiguous sentences. However, in this region of the unambiguous sentences, the full entropy will decrease to a certain degree, which entails some processing difficulties.

Lastly, under the surprisal hypothesis, it is predicted that RTs should be longer in the ambiguous region of the ambiguous than of the unambiguous sentences when main verbs have a high probability of SC subcategorization frame. Likewise, RTs should be longer in the disambiguating region of the ambiguous than of the unambiguous sentences when main verbs have a low probability of SC subcategorization frame.

4. Results: The analysis based on the PCFG from the L2 corpora

Figure 2 summarized the full entropy, entropy reduction, and surprisal estimated from the PCFG on the L2 corpora.[4] Interestingly, the full entropy and

4 When, for example, the verb claim takes either NP or SC in a ratio of 1:9, the full entropy is calculated by the following formula: h (the entropy before the verb) + 0.1 X H_{NP} (the internal entropy of NP) + 0.9 X H_{SC} (the internal entropy of SC).

surprisal show the patterns quite similar to those estimated from the PCFG on the L1 corpora (see Figure 2 and 3). However, the full entropy reduction was starkly different from its counterpart in the L1 study. In the L1 study, the value of the entropy reduction was the highest in the verb region regardless of sentence condition. The entropy reduction is higher in the verb region than in the disambiguating region.

In our L2 study, in the ambiguous sentence condition, the value of the entropy reduction is the highest in the ambiguous region; the entropy reduction is higher in the ambiguous than in the disambiguating regions. In the unambiguous sentence condition, the value of the entropy reduction is the highest in the disambiguating region; the entropy reduction is higher in the disambiguating than in the verb regions.

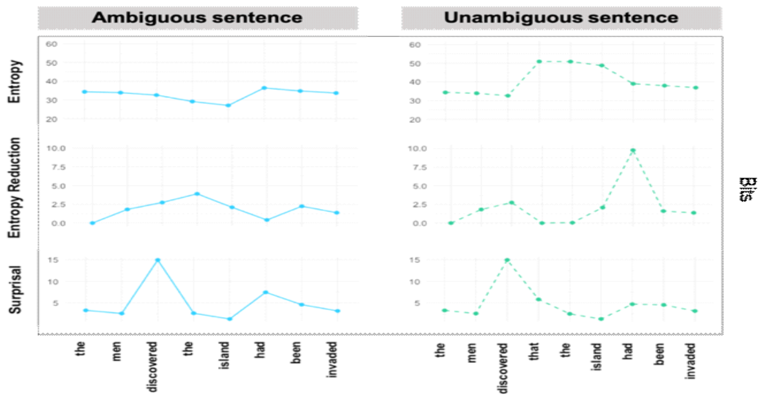

Figure 2. Word-by-word entropy, entropy reduction, and surprisal predictions in word-by-word from the probabilistic context-free grammar (PCFG) on the L2 corpora

---

Meanwhile, there are PCFG rules like: P(NP|accept) = 0.8; P(SC|accept) = 0.2, whereas P(S|forgot) = 0.09; P(PP|forgot) = 0.18; P(NP|forgot) = 0.55; P(Inf|forgot) = 0.14. The entropy of the verb *accept* is calculated by using the entropy formula: Entropy(accept) = $-(0.8 * \log_2(0.8) + 0.2 * \log_2(0.2)) = 0.721$. Likewise, the entropy of the verb *forgot* is calculated by using the entropy formula: Entropy(forgot) = $-(0.55 * \log_2(0.55) + 0.09 * \log_2(0.09)+0.18 * \log_2(0.18)+0.14 * \log_2(0.14)) = 1.629$.

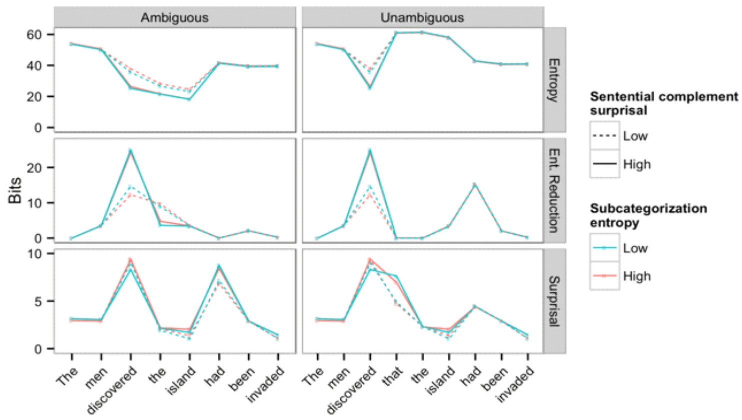

Figure 3. The information complexity metrics in word-by-word from Linzen and Jaeger (2016)

We suspect that the difference between the L1 and L2 corpora in regard to the locus of entropy reduction lies in the amount of information relating to lexical items, particularly main verbs. Since the L1 corpora grant ample information for verbs' subcategorization features, it follows that entropy reduction arose in the main verb region owing to less uncertainty about what will follow after the main verb. By contrast, the L2 corpora does not seem to provide sufficient information for such features. Thus, instead of being detected in the verb region, entropy reduction in the L2 study seems to have arisen in the post-verb, ambiguous or disambiguating region where more explicit syntactic information rather than more implicit lexical information plays a central role. In tandem with the estimates from the corpora that they are exposed to, L1 and L2 speakers are supposed to respond accordingly to the words in these regions, thus being reflected on RTs.

As already mentioned before, the competition hypothesis predicts a positive correlation between entropy and RTs, while the entropy reduction hypothesis predicts a positive correlation between entropy reduction and RTs. On the other hand, the Surprisal Hypothesis predicts a positive correlation between surprisal and RTs.

As in Table 1 below, the fitted model revealed a marginal significant effect of Sentence Condition in the verb region (β=-0.278, SE=0.160, t=-1.732, p=.083) and a significant main effect of Sentence Condition in the ambiguous (β=-0.375, *SE*=0.139, *t*=-2.693, *p*<.01) and in the disambiguating regions (β=-0.451, *SE*=0.178, *t*=2.529, *p*<.05). The participants' tendency to read differed across regions (Figure 4). Log RRTs were longer in the verb and the ambiguous regions of the ambiguous than of the unambiguous sentences. However, log RRTs were longer in the disambiguating region of the unambiguous than of the ambiguous sentences. A significant main effect of Entropy was found in the ambiguous region (β=0.155, *SE*=0.068, *t*=2.280, *p*<.05). Log RRTs were longer when the entropy increased. There was no entropy reduction effect in any of the regions.

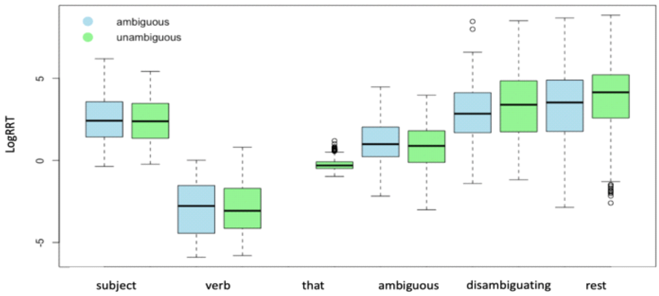

Figure 4. The mean log-transformed residual reading times (log RRTs) by sentence condition for each region[5]

There was a marginal significant effect of Surprisal in the verb region (β =-1.005, *SE*=0.517, *t*=-1.944, *p*=.061) and its significant main effect in the disambiguating region (β=3.456, *SE*=0.865, *t*=3.995, *p*<.001). When the surprisal increased, log RRTs were shorter in the verb region, but were longer in the disambiguating region. A marginal significant interaction between Sentence Condition and Surprisal was detected in the verb region (β=0.178, *SE*=0.102, *t*=1.744, *p*=.081). When the surprisal was low, the participants read faster in the unambiguous than in the ambiguous sentences. When the surprisal was high, the participants read faster in the ambiguous than in the unambiguous sentences. A significant main effect of Stimulus Order in all of the regions (Verb region: β =-0.042, *SE*=0.019, *t*=-2.134, *p*<.05; Ambiguous region : β=-0.091, *SE*=0.023,

5 Because the complementizer *that* was omitted in the case of an ambiguous sentence condition, there was no box plot at the region of *that*.

t=-3.837, p<.001; Disambiguating region : β=-0.130, SE=0.033, t=-3.917, p<.001), indicating that the participants read faster as the experiment made a progress. A significant main effect of Proficiency in the verb region was found (β=-0.104, SE=0.032, t=-3.193, p<.01). The participants as learners of high proficiency read relatively faster than participants as learners of low proficiency.

Table 1. Results of the statistical analyses

Model results. Time measure (Length-corrected log RTs)				
Verb region				
Coefficient	Estimate	SE	t-value	p-value
Intercept	-1.296	0.892	-1.453	.156
Sentence Condition (Ambiguous vs. Unambiguous)	-0.278	0.160	-1.732	.083†
Surprisal	-1.005	0.517	-1.944	.061†
Proficiency	-0.104	0.032	-3.193	< .01**
Stimulus Order	-0.042	0.019	-2.134	.033†
Sentence Condition X Surprisal	0.178	0.102	1.744	.081†
Proficiency X Stimulus Order	-0.072	0.019	-3.664	< .001***
Ambiguous region				
Coefficient	Estimate	SE	t-value	p-value
Intercept	-0.069	0.937	-0.074	.941
Sentence Condition	-0.375	0.139	-2.693	< .01**
Entropy	0.155	0.068	2.280	< .05*
Stimulus Order	-0.091	0.023	-3.837	< .001***
Disambiguating region				
Coefficient	Estimate	SE	t-value	p-value
Intercept	4.087	0.359	11.380	< .001***
Sentence Condition	0.451	0.178	2.529	< .05*
Surprisal	3.456	0.865	3.995	< .001***
Stimulus Order	-0.130	0.033	-3.917	< .001***

5. Discussion

This study has investigated whether Korean L2 learners of English predict the upcoming syntactic structure of sentence depending on the probability distribution of their expectations. Table 2 describes the predictions of the Surprisal Hypothesis as well as the competition and entropy reduction hypotheses for RTs, based on the PCFG-derived estimates.

Table 2. Predictions made by each hypothesis (based on the full entropy)

Hypotheses	Verb	Ambiguous	Disambiguating
Competition	Higher full entropy → Longer RTs ⊘	Longer RTs in unambiguous than in ambiguous sentences ✗	
Entropy Reduction	Higher full entropy, lower full entropy reduction → Shorter RTs ⊘	Shorter RTs in unambiguous than in ambiguous sentences ✓	Longer RTs in unambiguous than ambiguous sentences ✓
Surprisal			Shorter RTs in unambiguous than in ambiguous sentences ✗

Note: Predictions were derived by the competition and the entropy reduction hypotheses for each region of the target stimuli. The shaded cell means that both hypotheses do not predict any RT difference in the region. When predictions match with the results, the cell is marked with ✓; when predictions does not, the cell is marked with ✗; when predictions are not confirmed, the cell is marked with ⊘.

In our study, a significant effect of Entropy on RTs in the ambiguous region was detected: Higher entropy was correlated with longer RTs in the ambiguous region. In addition, a significant effect of Sentence Condition at all of the regions was found: RTs were longer in both the verb region and the ambiguous region of the ambiguous than of the unambiguous sentences, but were longer in the disambiguating region of the unambiguous than of the ambiguous sentences, supporting the entropy reduction hypothesis.

We take the results in this study to show that in accounting for Korean L2 learners of English's processing, the entropy reduction hypothesis is superior than the entropy hypothesis in fulfilling the predictions. Some studies claim that English native speakers continuously predict upcoming syntactic structure based on the probability distribution of their expectations during sentence processing. Linzen and Jaeger's (2016) L1 study in fact found a significant main effect of Entropy and Entropy Reduction in the main verb region: higher entropy was correlated with decreased RTs (β=-0.043, p<.001), and lower entropy reduction was also correlated with decreased RTs (β=0.042, p<.001) in this region. By contrast, Korean L2 learners of English seem not to have predicted the upcoming syntactic structure in the main verb region. There was no significant main effect of Entropy nor of Entropy Reduction in this region. Instead, a significant main effect of Entropy was observed in the ambiguous region. Thus, Korean L2 learners of English make use of more explicitly given syntactic information rather than more implicit lexical information like verbs' subcategorization frame. This claim is also supported by the results from the disambiguating region. Even when the L2 learners have already detected the syntactic structure in the unambiguous sentences after processing the complementizer *that*, they took longer time to process the disambiguating region of the unambiguous sentences than that of the ambiguous sentences. Recall that as the entropy reduction hypothesis suggests, the processing cost is affected by the reduction in uncertainty from word to word rather than the existence of uncertainty in the current region. Incidentally, under the Surprisal Hypothesis, shorter RTs are predicted in the disambiguating region of the unambiguous than in the ambiguous sentences, which is opposite to the prediction made by the entropy reduction hypothesis. Given that surprisal is positively correlated to RTs in the disambiguating region in Linzen and Jaeger's (2016) L1 study, it follows that surprisal may be more effective than entropy reduction in accounting for processing difficulty in this region. Unlike in their L1 study, however, in the disambiguating region of our L2 study, surprisal did not have the completely opposite effect from entropy reduction. Despite the strong surprisal effects in the disambiguating region, entropy reduction as well as surprisal may have a strong effect on processing difficulty. Indeed, both surprisal and entropy reduction had statistically significant effects on RTs, but the size of the surprisal effect was larger than the size of the entropy reduction effect (β = 8.925 vs. β = 1.525).[6]

Overall, Korean L2 learners of English differ from English native speakers in sentence processing. English native speakers predict the possible syntactic structure for the upcoming context promptly after reading a verb. By contrast, Korean L2 learners of English start to predict the upcoming syntactic structure in the post-verb, ambiguous and disambiguating regions when they are provided with more solid evidential sources such words and structures for syntactic

6 In Linzen and Jaeger's (2016) L1 study, in the disambiguating region, the size of the surprisal effect was also larger than the size of the entropy reduction effect, the former registering β = 0.06 and the latter β = 0.014, respectively. In linear regression with multiple independent variables such as surprisal and entropy reduction, the coefficient (β) tells you how much the dependent variable (i.e. RTs) is expected to increase when that independent variable increases by one, holding all the other independent variables constant.

analysis or reanalysis. This is a remarkable difference in sentence processing between L1 and L2 speakers. Still, the behavioral aspects of sentence processing by Korean L2 learners of English can be explained by the predictions that the entropy reduction and the Surprisal Hypotheses make. The uncertainty and unexpectedness (surprisal) play an instrumental role in shedding lights on the sentence processing by Korean L2 learners of English, especially at the ambiguous and the disambiguating regions. More broadly speaking, the approach integrating both computational and psycholinguistic methodologies provides new insights into L2 comprehension of syntactic structure during sentence processing.

References

Altmann, Gerry TM and Yuki Kamide. 1999. Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition* 73(3): 247-264.

Baayen, R. Harald and Peter Milin. 2010. Analyzing reaction times. *International Journal of Psychological Research* 3(2): 12–28.

Bar, Moshe. 2007. The proactive brain: using analogies and associations to generate predictions. *Trends in Cognitive Sciences* 11(7): 280-289.

Barr, Dale. J., Roger Levy, Christoph Scheepers and Harry J. Tily. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68(3): 255-278.

Bates, Douglas, Reinhold Kliegl, Shravan Vasishth, and Harald Baayen. 2015. Parsinomious mixed models. *arXiv preprint arXiv:1506.04967*

Bates, Douglas, Martin Maechler, and Ben Bolker. 2012. lme4: Linear mixed-effects models using S4 classes. R package version 0.999999-0.

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2014. Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823.*

Bell, Alan, Jason M. Brenier, Michelle Gregory, Cynthia Girand, and Dan Jurafsky. 2009. Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language* 60(1): 92-111.

Boston, Marisa Ferrara, John Hale, Reinhold Kliegl, Umesh Patil, and Shravan Vasishth. 2008. Parsing costs as predictors of reading difficulty: An evaluation using the Potsdam Sentence Corpus. *Journal of Eye Movement Research* 2(1): 1-12.

Chae, Eun-Young and Jeong-Ah Shin. 2015. A study of a timed cloze test for evaluating

L2 proficiency. *English Teaching* 70(3): 117-135.

Clark, Andy. 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences* 36: 181-204.

DeLong, Katherine A., Thomas P. Urbach, and Marta Kutas. 2005. Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience* 8(8): 1117–1121.

Demberg, Vera and Frank Keller. 2008. Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition* 109(2): 193-210.

Drummond, A. 2013. Ibex farm. Retrieved February 12, 2017.

Ehrlich, Susan F. and Keith Rayner. 1981. Contextual effects on word perception and eye movements during reading. *Journal of Memory and Language* 20(6): 641.

Elman, Jeffrey L., Mary Hare, and Ken McRae. 2004. Cues, constraints, and competition in sentence processing. In Michael Tomasello and Dan Slobin (eds.) *Beyond nature-nurture: Essays in honor of Elizabeth Bates*, 111-138. Mahwah, NJ: Lawrence Erlbaum Associates.

Federmeier, Kara D. 2007. Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology* 44(4): 491–505.

Ferreira, Fernanda and Charles Clifton. 1986. The independence of syntactic processing. *Journal of Memory and Language* 25(3): 348-368.

Frank, Stefan L. 2013. Uncertainty reduction as a measure of cognitive load in sentence comprehension. *Topics in Cognitive Science* 5: 475-494.

Frank, Stefan L., Leun J. Otten, Giulia Galli, and Gabriella Vigliocco. 2015. The ERP response to the amount of information conveyed by words in sentences. *Brain and Language* 140: 1-11.

Garnsey, Susan M., Neal J. Pearlmutter, Elizabeth Myers, and Melanie A. Lotocky. 1997. The contributions of verb bias and plausibility to the comprehension of temporarily ambiguous sentences. *Journal of Memory and Language* 37(1): 58-93.

Hale, John. 2001. A probabilistic Earley parser as a psycholinguistic model. In *Second Meeting of the North American Chapter of the Association for Computational Linguistics.*

Hale, John. 2003. The information conveyed by words in sentences. *Journal of Psycholinguistic Research* 32: 101-123.

Hale, John. 2006. Uncertainty about the rest of the sentence. *Cognitive Science* 30: 609-642.

Hale, John. 2011. What a rational parser would do. *Cognitive Science* 35: 399-443.

Henderson, John M., Wonil Choi, Steven G. Luke, and Joseph Schmidt. 2017. Neural correlates of individual differences in fixation duration during natural reading. *The Quarterly Journal of Experimental Psychology* 71(1): 314-323.

Huang, Bei-Bei, Guigang Zhang, and Phillip CY Sheu. 2008. A natural language database interface based on a probabilistic context free grammar. In *IEEE International workshop on Semantic Computing and Systems*, 155-162. IEEE.

Jennings, Fabio, Bill Randall, and L. K. Tyler. 1997. Graded effects of verb subcategory preferences on parsing: for constraint-satisfaction models. *Language and Cognitive Processes* 12(4): 485-504.

Kuznetsova, Alexandra, Per B. Brockhoff, and Rune HB Christensen. 2017. lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software* 82(13): 1-26.

Levy, Roger. 2005. Probabilistic models of word order and syntactic discontinuity. PhD Dissertation. Stanford University.

Levy, Roger. 2008. Expectation-based syntactic comprehension. *Cognition* 106(3): 1126–1177.

Levy, Roger and Edward Gibson. 2013. Surprisal, the PDC, and the primary locus of processing difficulty in relative clauses. *Frontiers in Psychology* 4: 229.

Linzen, Tal and T. Florian Jaeger. 2016. Uncertainty and expectation in sentence processing: Evidence from subcategorization distributions. *Cognitive Science* 40(6): 1382-1411.

Lowder, Matthew W., Wonil Choi, Fernanda Ferreira, and John M. Henderson. 2018. Lexical predictability during natural reading: Effects of surprisal and entropy reduction. *Cognitive Science* 42: 1166-1183.

Marcus, Mitchell, Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. Building a large annotated corpus of English: Penn Treebank. *Computational Linguistics* 19(2): 313-330.

McDonald, Scott A. and Richard C. Shillcock. 2003a. Eye movements reveal the on-line computation of lexical during reading. *Psychological Science* 14(6): 648-652.

McDonald, Scott A. and Richard C. Shillcock. 2003b. Low-level predictive inference in reading: The influence of probabilities on eye movements. *Vision Research* 43(16): 1735-1751.

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

O'Neill, Robert, Edwin T. Cornelius, and Gay N. Washburn. 1981. *American Kernel Lessons, AKL: Advanced student's book.* New York: Longman.

Roark, Brian, Asaf Bachrach, Carlos Cardenas, and Christophe Pallier. 2009. Deriving lexical and syntactic expectation-based measures for psycholinguistic modeling via incremental top-down parsing. In Philipp Koehn and Rada Mihalcea (eds.) *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing,* 324–333.

Shannon, Claude E. 1948. A mathematical theory of communication. *The Bell System Journal* 27(3): 379-423.

Shin, Jung A. 2019. A mixed-effects model-based analysis of experimental linguistic data: Focusing on the data from self-paced reading experiments (written in Korean). *English Linguistics* 19(1): 76-94.

Smith, Nathaniel J. and Roger Levy. 2013. The effect of word predictability on reading time is logarithmic. *Cognition* 128(3): 302–319.

Team, R. Core. 2018. R: A Language and Environment for Statistical Computing (Version 3.5. 2, R Foundation for Statistical Computing, Vienna, Austria, 2018).

Trueswell, John C., Michael K. Tanenhaus, and Christopher Kello. 1993. Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19(3): 528-553.

Trueswell, John C., Michael K. Tanenhaus, and Susan M. Garnsey. 1994. Semantic influences onparsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language* 33(3): 285-318.

Willems, Roel M., Stefan L. Frank, Annabel D. Nijhof, Peter Hagoort, and Antal Van den Bosch. 2016. Prediction during natural language comprehension. *Cerebral Cortex* 26(6): 2506-2516.

Wu, Stephen, Asaf Bachrach, Carlos Cardenas, and William Schuler. 2010. Complexity metrics in an incremental right-corner parser. In Jan Hajič, Sandra Carberry, Stephen Clark, Joakim Nivre (eds.) *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics,* 1189-1198.

Yuasa, Katsura. 2010. English Textbooks in Japan and Korea. *Journal of Pan-Pacific Association of Applied Linguistics* 14(1): 147-158.

Yun, Jiwon, Zhong Chen, Tim Hunter, John Whitman, and John Hale. 2015. Uncertainty in processing relative clauses across East Asian languages. *Journal of East Asian Linguistics* 24(2): 113-148.

Appendix A: Stimuli

Materials	
1	The men discovered (that) the island had been invaded by the enemy.
2	The women revealed (that) the secret had been exposed by the officials.
3	The man noticed (that) the mistake had not happened due to negligence.
4	The woman assumed (that) the blame might have belonged to the driver.
5	They all indicated (that) the problem might not bother the entire team.
6	Two people found (that) the equipment should be reported stolen right away.
7	Some people sensed (that) the conflict should be resolved quickly and peacefully.
8	The woman determined (that) the estimate had been inflated by the accountant.
9	Two people heard (that) the album had been criticized in the magazine.
10	Some people understood (that) the message had not meant much to foreigners.
11	They all read (that) the newspaper might be going out of business.
12	The women worried (that) the parents might have become quite restless recently.
13	The man taught (that) the children should be sheltered from all harm.
14	The men projected (that) the film would not gross enough in cinemas.
15	They all claimed (that) the luggage had been stolen from the hotel.
16	Some people regretted (that) the decision had been reached without any discussion.
17	The men remembered (that) the appointment had not changed since last week.
18	The women warned (that) the drivers might have drunk too much vodka.
19	Many people feared (that) the future might not hold hope for them.
20	The man proposed (that) the idea should be abandoned for financial reasons.
21	Two people suggested (that) the scene should be filmed right before sunset.
22	The woman announced (that) the wedding would be postponed until late August.
23	The men forgot (that) the details had been worked out in advance.
24	The man observed (that) the patient had been sent home too early.
25	The woman recalled (that) the speech had not gone over very well.
26	The women answered (that) the questions might be discussed during the meeting.
27	Some people added (that) the numbers might have decreased since last year.
28	Two people wrote (that) the interview should be conducted over the phone.
29	Many people advised (that) the president should be considering further budget cuts.
30	The men begged (that) the judge would not treat the defendant harshly.

Appendix B

Model estimates of log-transformed residual reading times at verb, ambiguous, and disambiguating regions of target sentences. Models included centered fixed effects of Sentence Condition (Ambiguous vs. Unambiguous), Surprisal (the value of surprisal), Entropy (the value of entropy), Language Proficiency (the score of cloze test), Stimulus Order (Trial position of the stimuli), and all interactions between Sentence Condition, Surprisal, Entropy, Language Proficiency, and Stimulus Order. Random intercepts of participants and items were included, and random slopes were included as model convergence allowed.

Model results. Time measure (Length-corrected log RTs)				
Verb region				
Coefficient	Estimate	SE	t-value	p-value
Intercept	-1.531	1.430	-1.071	0.292
Sentence Condition (Ambiguous vs. Unambiguous)	-0.236	0.241	-0.977	.329
Surprisal	-0.901	0.727	-1.29	.225
Entropy	-0.518	2.145	-0.242	.810
Proficiency	-0.107	0.032	-3.278	< .01**
Stimulus Order	-0.041	0.019	-2.098	< .05*
Sentence Condition X Surprisal	0.141	0.122	1.156	.248
Sentence Condition X Entropy	0.098	0.372	0.263	.792
Surprisal X Entropy	0.226	1.053	0.215	.831
Proficiency X Stimulus Order	-0.071	0.019	-3.625	< .001***
Sentence Condition X Surprisal X Entropy	-0.090	0.200	-0.452	.651
Ambiguous region				
Coefficient	Estimate	SE	t-value	p-value
Intercept	-0.069	0.937	-0.074	.941
Sentence Condition	-0.375	0.139	-2.693	< .01**
Entropy	0.155	0.068	2.280	< .05*
Surprisal	-1.285	1.209	-1.064	.288
Proficiency	-0.036	0.046	-0.779	.442
Stimulus Order	-0.091	0.023	-3.837	< .001***

Sentence Condition X Entropy	-0.117	0.113	-1.027	.304
Disambiguating region				
Coefficient	Estimate	SE	t-value	p-value
Intercept	4.087	0.359	11.380	< .001***
Sentence Condition	0.451	0.178	2.529	< .05*
Entropy	-0.049	0.144	-0.343	.731
Surprisal	3.456	0.865	3.995	< .001***
Proficiency	-0.060	0.072	-0.845	.405
Stimulus Order	-0.130	0.033	-3.917	< .001***
Sentence Condition X Entropy	0.082	0.250	0.330	.741

Significance levels: †p < .1, *p< .05, **p< .01, ***p<.001

Appendix C

Model estimates of log-transformed residual reading times at Verb, Ambiguous, and Disambiguating regions of target sentences. Models included centered fixed effects of Sentence Condition (Ambiguous vs. Unambiguous), Surprisal (the value of surprisal), Entropy Reduction (the value of entropy reduction), Language Proficiency (the score of cloze test), Stimulus Order (Trial position of the stimuli), and all interactions between Sentence Condition, Surprisal, Entropy Reduction, Language Proficiency, and Stimulus Order. Random intercepts of participants and items were included, and random slopes were included as model convergence allowed.

Model results. Time measure (Length-corrected log RTs)				
Verb region				
Coefficient	Estimate	SE	t-value	p-value
Intercept	-1.296	0.892	-1.453	.156
Sentence Condition (Ambiguous vs. Unambiguous)	-0.278	0.160	-1.732	.083†
Surprisal	-1.005	0.517	-1.944	.061†
Entropy reduction	0.071	0.922	0.078	.938
Proficiency	-0.104	0.032	-3.193	< .01**
Stimulus Order	-0.042	0.019	-2.134	< .05*
Sentence Condition X Surprisal	0.178	0.102	1.744	.081†
Sentence Condition X Entropy reduction	-0.196	0.218	-0.900	0.368

Surprisal X Entropy reduction	0.062	0.499	0.125	.901
Proficiency X Stimulus Order	-0.072	0.019	-3.664	< .001***
Sentence Condition X Surprisal X Entropy Reduction	0.134	0.146	0.919	.358
Ambiguous region				
Coefficient	Estimate	SE	t-value	p-value
Intercept	0.830	0.883	0.940	.348
Sentence Condition	-0.130	0.079	-1.638	.102
Entropy Reduction	-0.047	0.057	-0.820	.413
Proficiency	-0.025	0.046	-0.550	.587
Stimulus Order	-0.095	0.024	-3.958	< .001***
Disambiguating region				
Coefficient	Estimate	SE	t-value	p-value
Intercept	5.097	0.774	6.785	< .001***
Sentence Condition	1.525	1.194	1.277	.202
Surprisal	8.925	3.447	2.589	< .01**
Entropy Reduction	-0.243	0.404	-0.603	.546
Proficiency	-0.057	0.072	-0.797	.432
Stimulus Order	-0.130	0.033	-3.923	< .001***
Sentence Condition X Surprisal	8.077	5.538	1.459	.145
Sentence Condition X Entropy Reduction	-1.179	0.695	-1.697	.090†
Surprisal X Entropy Reduction	-2.297	1.730	-1.327	.185
Proficiency X Stimulus Order	0.015	0.032	0.465	.642
Sentence Condition X Surprisal X Entropy Reduction	-5.128	3.077	-1.666	.096†

Significance levels: †p < .1, *p< .05, **p< .01, ***p<.001

Euhee Kim
Associate Professor
Department of English
Computer Science and Engineering, School of IT Convergence Engineering,
Shinhan University, 30, Byeolmadeul-ro 40-gil,
Dongducheon City, Gyeonggi Province, Korea, 483-777
E-mail: euhkim@shinhan.ac.kr

Myung-Kwan Park
Professor
Department of English

Dongguk University
30, 1-gil, Pildong-ro, Chung-gu
Seoul, 04620, Korea
E-mail: parkmk@dgu.edu

Hye-Jin Seo
Ph.D Student
Department of English
Dongguk University
30, 1-gil, Pildong-ro, Chung-gu
Seoul, 04620, Korea
E-mail: seohj0951@gmail.com

Received: 2020. 04. 20.
Revised: 2020. 08. 24.
Accepted: 2020. 08. 29.