A Test of Frequency-based Processing:
Evidence from Korean Numeral Classifier Constructions[*]

On-Soon Lee
(Korea University)

The Korean numeral classifier constructions are a good choice to test frequency-based processing because one of the two types of construction (prenominal and postnominal) occurs more often than the other in production, and distributional frequency matters in comprehension. The study conducted an experiment with a self-paced reading task. Findings showed that (i) comprehension accuracy is significantly higher for the more-frequent postnominal constructions than for the less-frequent prenominal constructions and (ii) reading times at the critical region of prenominal constructions are faster in the condition with a general classifier than in the condition with a specific classifier, although reading times in postnominal constructions did not differ in the two conditions. These findings indicate that the familiarity (i.e., experience) of frequent structures facilitates faster and more accurate comprehension. The study thus contributes crosslinguistic evidence to add to the evidence supporting the frequency-based account based on distribution.

Keywords: sentence processing, frequency, distribution, syntactic parsing, Korean

1. Introduction

Recent psycholinguistic research has been interested in frequency-based theories,[1] which propose that the key constraining factor in production

1) 'Expectation' is broadly used to refer to predictability or frequency, but in this study, 'expectation' refers to the conditional probability of a word occurring in a given context, whereas 'distribution' and 'frequency' refer to the occurrences of structures in the corpus.

and comprehension is not memory but experience; in other words, the structures with which readers have more direct experience might be easier to comprehend. So-called word-order frequency-based theories (see Levy, Fedorenko, & Gibson, 2013 for a review) propose that the surface ordering of words that occurs more frequently in the input would be preferred and thus easier to process during comprehension (MacDonald & Christiansen, 2002, 2009; Wells, Christiansen, Race, Acheson, & MacDonald, 2009). The key idea is that comprehending rarely occurring structures, of which comprehenders have less experience, is more difficult than comprehending relatively more frequent structures. In such theories, the greater processing difficulty of object relative clauses (e.g., *the man who the woman liked*⋯) than subject relative clauses (e.g., *the man who liked the woman*⋯) is due to the frequency of the patterns of the two types of relative clauses' surface ordering in English; the object relatives' surface pattern (Object — Subject — Verb) occurs less often than that of the subject relatives (Subject — Object — Verb). Hale (2001) and Levy (2008) have clarified how syntactic expectation (based on a word's probability of occurring given the context, *surprisal*), which emerges from comprehenders' experience of distribution, facilitates the comprehending of sentences. However, whether the distributional frequency theory (MacDonald & Christiansen, 2002, 2009; Wells, Christiansen, Race, Acheson, & MacDonald, 2009) is applicable to the case of Korean numeral classifier constructions has rarely been tested. Therefore, this study adds to crosslinguistic research on frequency-based processing by examining the effect of distributional frequency in the comprehension of Korean numeral classifier constructions, focusing on two different types of constructions (prenominal and postnominal) and two different types of classifiers (specific and general).

The paper is organized as follows. Section 2 introduces findings from previous studies. Sections 3 and 4 describe the current study and its findings. Section 5 discusses the implications of the experimental results, and Section 6 concludes the paper.

2. Background

2.1. Memory, expectation, and frequency in language processing

In the psycholinguistic literature, several previous studies have focused on memory-based accounts of sentence comprehension (Clifton & Frazier, 1989; Frazier, 1979; Frazier & Fodor, 1978; Gibson, 1998, 2000; Just & Carpenter, 1992; Lewis & Vasishth, 2005; Lewis, Vasishth, & Van Dyke, 2006). A well-established example of a memory-based account is dependency locality theory (DLT) (Gibson, 1998, 2000). According to DLT, processing difficulty is relevant to two aspects of language comprehension: (i) storage cost and (ii) integration cost. Gibson defines storage cost in terms of the number of syntactic heads that are necessary to form a grammatical sentence. Integration cost is defined in terms of the number of new discourse referents (i.e., referential elements) intervening between an incoming word and the syntactic head with which the incoming word is associated. For example, Gibson and his colleagues suggested that the greater number of intervening words between a head noun and a gap in direct object relative clauses makes them more difficult to comprehend than subject relative clauses.

However, the work of Hale (2001) and Levy (2008) has pointed to the role of expectation (i.e., what can be expected based on the conditional probability of a given word appearing in a given context) in sentence comprehension. In particular, Levy proposed the *surprisal* hypothesis, which suggests that the different degrees of expectation for an upcoming word triggered by a given context lead to different degrees of processing difficulty. Several online comprehension experiments have been conducted in various languages to test this expectation-based account (Chinese: Jäger, Chen, Li, Lin, & Vasishth, 2015 and Wu, Kaiser, & Anderson, 2009; Russian: Levy, Fedorenko, & Gibson, 2013; German: Levy & Keller, 2013; English: Linzen & Jaeger, in press and Staub, 2010; Korean: Yun & Hong, 2014 and Yun, Nam, Yoo, & Hong, 2015). For example, Staub (2010) compared the expectation-based account and the memory-based account in an eye-tracking study on processing object relative clauses with twenty-eight English native speakers as participants. He assumed that the

memory-based account (Gibson, 1998; Gordner & Gibson, 2005) predicts difficulty at the matrix verb (e.g., *hurried* in [1b]), whereas the expectation-based account predicts difficulty earlier (e.g., *the* or *fireman* in [1b]).

(1) a. The employees that noticed the fireman hurried across the open field.

b. The employees that the fireman noticed hurried across the open field.

Staub found that the mean *go-past* time on noun phrases (e.g., *fireman* in [1b]) — the sum of all fixations beginning with the first fixation on a critical region (e.g., *that, the, fireman, noticed*, and *hurried* in [1b]) until the reader leaves the critical regions — was significantly longer for object relative clauses than subject relative clauses. The detailed analysis indicated that the readers had difficulty processing object relative clauses earlier, such as at *fireman* in (1b). The study's results are consistent with the prediction of an expectation-based account; the difficulty of processing object relative clauses appears as soon as a reader encounters a relative pronoun, indicating that the predictability triggered at the relative pronoun affects the reader's expectations for an upcoming word.

A recent study by Jäger et al. (2015) tested the role of expectation in the processing of relative clauses in Chinese with a self-paced reading task with forty-nine Chinese native speakers as participants. As (2) and (3) show, Chinese relative clauses are prenominal — the head noun (e.g., *girl* in [2] and [3]) appears after the relative clause. For Chinese, a memory-based account predicts that object relative clauses (2b and 3b) should be easier to comprehend than subject relative clauses (2a and 3a) because the distance between the head noun and the gap inside the relative clause is greater in subject relative clauses than in object relative clauses. In contrast, the expectation-based account based on the conditional probability of head nouns predicts longer reading times for the object relative clauses (i.e., a subject relative advantage).

(2) a. subject-modifying subject RC[2)]

哪个 　　　上个月 　　　邀清了 　　男孩 的 　女孩 　知道 王老師
Na-ge 　t_i shanggeyue 　yaoqing-le 　nanhai de 　nühai $_i$ renshi laoshi
DET-CL[RC t_i 　last month 　invite-ASP 　boy REL] 　girl$_i$ 　know teacher
'The girl who invited the boy last month knows the teacher.'

b. subject modifying object RC

哪个 　　　上个月 　　男孩 邀清了 　的 　女孩 　知道 王老師
Na-ge 　shanggeyue nanhai yaoqing-le 　t_i de 　nühai $_i$ renshi laoshi
DET-CL [RC last month 　boy 　invite-ASP 　t_i REL] girl$_i$ know teacher
'The girl who the boy invited last month knows the teacher.'

(3) a. object-modifying subject RC

王老師 　　　知道 哪个 　　　　上个月 　　　邀清了 男孩 的 女孩
Wang laoshi renshi na-ge 　　t_i shanggeyue 　yaoqing-le nanhai de nühai $_i$
teacher know DET-CL [RC t_i last month 　invite-ASP boy 　REL] 　girl$_i$
'The teacher knows the girl who invited the boy last month.'

b. object-modifying object RC

王老師 　　　知道 哪个 　　　上个月 　　　男孩 邀清了 　的 女孩
Wang laoshi renshi na-ge 　　shanggeyue 　nanhai yaoqing-le t_i de nühai $_i$
teacher know DET-CL[RC last month 　boy 　invite-ASP t_i 　REL] girl$_i$
'The teacher knows the girl who the boy invited last month.'

To confirm the conditional probability of the head noun in Chinese relative clauses, the study first conducted a corpus analysis using Chinese Treebank 7.0 and then conducted an experiment with a sentence completion task. They found that (i) subject relative clauses as in (2a) and (3a) frequently occur and (ii) the participants preferred to complete sentences by producing patterns such as (2a) and (3a). The experiment showed the effect of the conditional probability of an overt head noun appearing in a given Det + Cl + Adv + subject RC/object RC sequence; the predictability of a word in context affected the ease with which the parsers built the structure, leading to easier processing of subject RCs. In addition, as predicted based on the estimated predictability of the head nouns, the findings from the self-paced reading task showed that reading times at the critical region in subject relative clauses (e.g., *boy* in [2a]) were significantly faster than those at the critical region in object relative clauses (e.g., *invite* in [2b]), and this pattern held for both subject-modifying (e.g., [2a-b]) and ob-

---

2) ADV = adverb; ASP = aspect; DET = determiner; CL = classifier; RC = relative clause; REL = relative

ject-modifying (e.g., [3a-b]) relative clauses. Moreover, reading times at the head noun (e.g., *girl* in [2] and [3]) did not differ. These findings are consistent with the prediction of the expectation-based account, but counter to the prediction of the memory-based account.

However, the estimate of the conditional probability of a word may not fully explain the advantage of subject relative clauses in Chinese; some studies have argued that processing object relative clauses is easier than processing subject relative clauses in Chinese (Hsiao & Gibson, 2003; Gibson & Wu, 2013). If so, frequency-based theories might suggest a possible constraining factor on processing: Chinese is a SVO language, so if distributional frequency-based processing applies to sentences like those in (2) and (3), the object relative clauses in (3) may be preferred over the subject relative clauses in (2) due to their SVO-like word order. Following this view, MacDonald and her colleagues proposed that learners' experience of frequency occurrence is a key constraining factor in processing sentences (MacDonald & Christiansen, 2002; Reali & Christiansen, 2007a, 2007b; Gennari & MacDonald, 2009), so a highly frequent structure is easier to process. For example, Reali and Christiansen (2007a) found that in English, pronominal object relative clauses (4a) are easier to process only when a personal pronoun is in the embedded position. This finding is consistent with the findings of corpus analysis: Pronominal object relative clauses with a personal pronoun in embedded position occur more frequently than pronominal subject relative clauses. Furthermore, they suggested that statistical (i.e., frequency or distributional) information should be taken into consideration in accounts of sentence processing.

(4) a. The lady that I visited enjoyed the meal.
    b. The lady that visited me enjoyed the meal.

A recent work by Wells, Christiansen, Race, Acheson, and MacDonald (2009) emphasized the effect of distributional frequency on sentence comprehension. For example, sentences like the one in (5a), with a subject relative clause, are more frequent than sentences like the one in (5b),

with a direct object relative clause; therefore, processing direct object relative clauses is more difficult than processing subject relative clauses.

(5) a. The man that likes the woman is tall.
    b. The man that the woman likes is tall.

More specifically, the authors claim that the difficulty of processing direct object relative clauses is due to their less frequent word order; this is called the word-order frequency theory. Subject relative clauses display a word order similar to that of simple transitive sentences (e.g., *The man likes the woman*.); that is, an SVO-like word order. On the other hand, direct object relative clauses manifest an OSV-like word order (e.g., *The man, the woman likes.*). In other words, because readers frequently encounter the canonical word order in English (i.e., SVO), the subject relative clauses that manifest an SVO-like word order are easier to comprehend.

Based on this hypothesis, Wells et al. (2009) assumed that exposing readers to more direct object relative clauses would decrease the difficulty of processing such clauses. In their study, 97 native speakers of English completed a pre-test, two input sessions, and a post-test. For the pre-test and post-test, all the participants completed a self-paced reading task including subject and direct object relative clauses with fillers. For the two input sessions, the participants were randomly divided into two groups; one read materials with many subject and direct object relative clauses, while the other read materials with only simple transitive sentences. The group that received the relative clause input demonstrated a decrease in reading times that was greater for direct object relative clauses than for subject relative clauses, whereas the group that received simple transitive sentence input did not. The group receiving relative clause input also read significantly faster at the critical region of direct object relative clauses (e.g., *is* in [5]) in the post-test than in the pre-test. These findings imply that (i) the infrequent structure of direct object relative clauses causes greater difficulty in their comprehension and (ii) frequency in input could improve learners' comprehension. These findings are confirmed in the replication of that study (O-S Lee, 2014).

The findings from the previous studies thus shed light on another factor (i.e., distribution or frequency) that affects sentence comprehension, but questions still remain. Further investigation is necessary to examine how generally the frequency-based account applies crosslinguistically. This study takes a step toward addressing this gap by testing the expect- ation-based account in an experiment on the comprehension of Korean numeral classifier constructions, which are introduced in the following section.

2.2. Properties of Korean numeral classifier constructions

Korean numeral classifier constructions provide a useful case for testing the predictions of the frequency-based account. This section introduces the semantic and syntactic properties of Korean numeral classifier constructions. From the semantic perspective, Korean classifiers can be divided into two types according to the degree of semantic relatedness between the classifiers and associated nouns: specific and general (Sohn, 1999; Unterbeck, 1994). The sentences in (6) provide examples.

(6) a. Specific classifier for books, *kwen*[3]

    haksayng-i   twu **kwen**-uy   **chayk-ul**   sass-ta.
    student-nom   two $CL_{specific}$-gen book-acc   bought-decl
    'A student bought two books.'

  b. General classifier, *kay*[4]

    haksayng-i   twu **kay**-uy   **chayk-ul**   sass-ta.
    student-nom   two $CL_{general}$-gen   book-acc   bought-decl
    'A student bought two books.'

  c. Incorrect use of specific classifier for clothing, *pel*

    *haksayng-i   twu **pel**-uy   **chayk-ul**   sass-ta.
    student-nomtwo $CL_{specific}$-gen   book-acc   bought-decl
    'A student bought two books.'

---

3) NOM = nominative; GEN = genitive; CL = classifier; ACC = accusative; DECL = declarative

4) A reviewer pointed out the unnaturalness of (6b), indicating that *kay* might not replace all specific classifiers for inanimate nouns. However, in an off-line naturalness judg- ment task, the general classifier *kay* with the noun *chayk* 'book' seemed to be relatively natural to participants; its average rating was 3 (out of 5), whereas the average rating of the specific classifier and noun was 3.4 (out of 5), indicating that both were accept- able to participants.

As seen in (6a), *kwen* (a classifier for books) is a specific classifier, which can be used only for the semantically associated noun (*book*). As seen in (6b), *kay* (a classifier for inanimate entities) is a general classifier that can appear as a substitute for the specific classifier. Semantic disagreement between a classifier and noun, as seen in (6c), where *pel* (a classifier for clothing) occurs with the noun *chayk* 'book', results in ungrammaticality.

From the syntactic perspective, there are two types of Korean numeral classifier constructions depending on the position of the numeral classifier: prenominal constructions and postnominal constructions. In the example in (7a), a numeral classifier (*twu-kwen* 'two-CL') with the genitive marker *-uy* appears before the associated noun (*chayk* 'book'); this is called the prenominal construction. In the example in (7b), the numeral classifier (*twu kwen* 'two-CL'), with no genitive marking follows the associated noun (*chayk* 'book'); this is called the postnominal construction (H Sohn, 1999).[5)]

> (7) a. haksayng-i   twu **kwen**-uy   **chayk-ul**   sass-ta.
>   student-nom   two **CL**$_{specific}$-gen   book-acc   bought-decl
>   'A student bought two books.'
>   b. haksayng-i   **chayk-ul**   **twu kwen**   sass-ta.
>   student-nom   book-acc   two **CL**$_{specific}$ bought-decl
>   'A student bought two books.'

In (7), the numeral classifier directly modifies the associated noun *chayk* ('book'), whereas in (7b), the numeral classifier *twu kwen* ('two') is floated—that is, it appears after the associated noun (*book*). Focusing on this difference in structures, a few researchers have sought a syntactic approach that could explain all possible constructions with numeral classifiers by using psycholinguistic methodologies (e.g., self-paced reading times and event-related potential [ERP]). For example, H K and E Oh (2012) examined the validity of a hybrid approach to explain several patterns of floating

---

5) Previous studies on Korean numeral classifier constructions have categorized several types, based on different syntactic perspectives (see H Ko & E Oh, 2012 for a literature review). Without making a claim regarding the categorization of these structures, this study focuses on the two types of constructions including object-oriented numeral classifiers described here.

numeral classifiers, whereas M-K Park et al. (2014) tested the stranding view of floating numeral classifiers in Korean. There remain many unresolved theoretical issues about how prenominal constructions differ structurally from postnominal constructions, but these matters are beyond the scope of this paper (see H Ko, 2007; H Ko & E Oh, 2012; M-K Park, W Chung, & K Choi, 2014 for a summary of the literature).

3. Corpus analysis

It is assumed that the general classifier used for counting inanimate entities frequently appears in spoken language, particularly in postnominal constructions. Thus, the corpus analysis was conducted in order to ascertain the distributional frequency of Korean numeral classifiers before creating the materials for the online self-paced reading task. The Sejong Corpus, which was used for this analysis, includes both written and spoken language, but this study is interested in the distributional frequency of production, so the analysis was conducted only on the spoken language corpus (2,050,000 words), which consists of the scripts of TV dramas and news. MonoConc Pro (version 2.2) was used to extract a total of 574 sentences with general and specific classifiers whose associated nouns were in the direct object position.[6] Of the 574 sentences, 122 contained prenominal constructions including only general classifiers and 452 contained postnominal constructions with both general and specific classifiers, as seen in Figure 1 (see Appendix A for details). These results are similar to those of J-B Kim and J Yang's (2006) corpus study; postnominal constructions occur more often than prenominal constructions (e.g., in 86 vs. 104 out of 694 sentences from the Sejong Treebank Corpus in J-B Kim & J Yang, 2006).

---

6) Since 1998, a national project has aimed at building an information technology system for the Korean language. This has produced the Sejong Corpus, the largest corpus of Korean language data as of 2004 (Bley-Vroman & H Ko, 2005, p. 258).
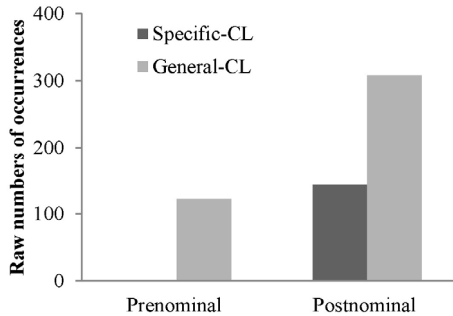
Figure 1. The distributional frequency of Korean numeral classifiers in spoken data from the Sejong Corpus.

As seen in Figure 1, classifiers appear more often in postnominal constructions (144 occurrences of specific classifiers and 308 occurrences of general classifiers) than in prenominal constructions (122 occurrences of only general classifiers) in spoken texts. The results from the corpus analysis show that only general classifiers at the direct object position appear in prenominal constructions. In postnominal constructions, general classifiers appear more frequently than specific classifiers. Based on these frequency patterns of the two types of numeral classifiers, three research questions were developed to test the predictions of the frequency-based account for the results of the self-paced reading experiment:

1. Do participants comprehend postnominal constructions more easily than prenominal constructions? It is predicted that the percentage of correct responses to comprehension questions will be higher in postnominal constructions than in prenominal constructions.[7]
2. Will the prenominal constructions with general classifiers be comprehended more easily than the prenominal constructions with specific classifiers? It is predicted that reading times at the associated nouns following general classifiers will be relatively faster than at the associated nouns following specific classifiers.

---

7) The mean accuracy rate of responses to comprehension questions is considered a dependent variable because specific or general classifiers might cause temporary ambiguity related to the classifiers' semantic agreement with associated nouns.

3. Will general classifiers in postnominal constructions be compre-
hended more easily than specific classifiers? It is predicted that read-
ing times at the general classifiers will be relatively faster than reading
times at the specific classifiers.

4. Current Study

4.1. Participants

Forty-three native speakers of Korean (29 women and 14 men; mean
age = 22.3, SD = 2.1, range = 19-27), all of them undergraduates, partici-
pated in the self-paced reading experiment. They were paid five dollars
each for their participation in the experiment.

4.2. Materials

The experiment used two types of classifiers (specific vs. general) ×
two types of constructions (prenominal vs. postnominal). Four lists in
a Latin-square design were created to present to participants. Tables 1
and 2 show one set of the four conditions used in this experiment. The
results of a pilot study showed no difference in reading times for pre-
nominal and postnominal constructions for sentences in which the classi-
fiers were closely adjacent to the associated nouns. This indicates the
possibility that readers do not have enough time to understand semantic
feature agreement between the classifier and the noun in such sentences;
that is, readers can easily ignore agreement in real time. Moreover, the
critical region of interest is region 5 (*pencil*), but the spill-over effect occurs
before and after the critical region: The intervening words are needed.
Therefore, the test sentences manipulated the distance between the classi-
fiers and associated nouns across the two constructions by including inter-
vening adjectives.[8]

---

8) A reviewer expressed the concern that the frequency rate of intervening words in pre-
   nominal and postnominal constructions might influence comprehension of the follow-
   ing classifiers. Although the intervening adjectives are needed for methodological rea-

Table 1. A Sample Set of Experimental Conditions with Prenominal Constructions[9]

Conditions	Regions						
	1	2	3	4	5	6	7
Specific-CL	sonyen-i	ecey	**sey clawu-uy**	kitalan nolan-sayk	**yenphil-ul**	mwunpangkwu-eyse	sass-ta
	boy-NOM	yesterday	three CLspecific-GEN	long,yellow-color	pencil-ACC	stationery store-LOC	bought
General CL	sonyen-i	ecey	**sey kay-uy**	kitalan nolan-sayk	**yenphil-ul**	mwunpangkwu-eyse	sass-ta
	boy-NOM	yesterday	three CLgeneral-GEN	long,yellow-color	pencil-ACC	stationery store-LOC	bought

'A boy bought three long, yellow pencils at a stationery store yesterday.'

As shown in Table 1, while the direct object (region 5) remains the same in the two classifier conditions, the numeral classifiers (region 3) vary in the degree of their semantic relatedness to the direct object. In Table 2, the choice of numeral classifiers and associated nouns is the same as in Table 1, but the associated nouns occur before the numeral classifiers.

Table 2. A Sample Set of Experimental Conditions with Postnominal Constructions

Conditions	Regions						
	1	2	3	4	5	6	7
Specific-CL	sonyen-i	kitalan nolan-sayk	**yenphil-ul**	ciwukay-wa hamkkey	**sey calwu**	mwunpangkwu-eyse	sass-ta
	boy-NOM	long,yellow-color	pencil-ACC	eraser-and together	three CLspecific	stationery store-LOC	bought
General CL	sonyen-i	kitalan nolan-sayk	**yenphil-ul**	ciwukay-wa hamkkey	**sey calwu**	mwunpangkwu-eyse	sass-ta
	boy-NOM	long,yellow-color	pencil-ACC	eraser-and together	three CLgeneral	stationery store-LOC	bought

'A boy bought three long, yellow pencils with erasers at a stationery store.'

It is unclear whether agreement between nouns and classifiers is processed during online comprehension. As Table 2 shows, sentences with the postnominal construction included local nouns (e.g., *eraser*) intervening between nouns and classifiers, in order to address the spill-over effect. Another reason for this manipulation is that it is assumed to be harder to check for semantic agreement between classifiers and associated nouns in postnominal constructions (i.e., when the noun precedes the classifier) than in prenominal constructions (i.e., when the noun follows). The local

---

sons in this study, I agree that this is a concern that must be addressed in further research.

9) LOC = locative

nouns (e.g., *eraser*) inserted in the postnominal constructions belong to the same semantic group as the preceding associated nouns (e.g., 'pencil'), but do not match categorically with the following specific classifiers, in order to facilitate the readers' ability to check semantic agreement of the classifiers with the preceding associated nouns.

   In this experiment, twelve different types of specific classifiers and one general classifier, *kay*, were employed; each specific classifier appears twice in each list, while the general classifier appears twelve times.[10] Each list includes twenty-four target sentences with forty-eight filler sentences. Filler sentences do not include any numeral classifiers.

4.3. Procedure

   The experiment was conducted on a computer running E-prime (Version 2.0). Each sentence was presented one region at a time on the computer screen, left to right, in a noncumulative, moving-window manner as a participant pushed the space bar (Just, Carpenter, & Woolley, 1982). Each participant completed the self-paced reading task in one session lasting less than 30 minutes. Participants were asked to sit in front of the computer screen. Then, they were asked to read a total of 72 Korean sentences one at a time, region by region, as they pressed the space bar; after each sentence, they were asked to give a yes/no answer to a comprehension question such as (8b). All reading times for each region were recorded, along with responses to comprehension questions.

(8) a. sonyen-i   **sey calwu-uy**   kitalan nolan-sayk **yenphil-ul** mwunpangkwu-eyse sass-ta.
       boy-nom three CL$_{specific}$-gen long,yellow-color pencil-acc stationery store-loc bought.
       'A boy bought three long, yellow pencils with erasers at a stationery store.'

   b. sonyen-i yenphil-ul sey calwu sass-ni?
      'Did a boy buy three pencils?'
      Correct Answer: Yes

---

10) All specific classifiers are used as counters for inanimate entities.

4.4. Predictions

The distributional frequency of numeral classifier constructions in Korean leads to the following predictions. First, the distribution from the corpus analysis showed that postnominal constructions occur more often than prenominal constructions; therefore, the percentage of correct responses to the comprehension questions in the postnominal constructions, regardless of the type of classifier, will be significantly higher than the percentage of correct responses in the prenominal constructions.

Second, in the prenominal constructions, the greater frequency of general classifiers will facilitate the comprehension of sentences with general classifiers, leading to faster reading times at the critical regions. As seen in Table 1, the critical region in prenominal constructions is region 5. When participants encounter an associated noun (e.g., *pencil*) after passing the numeral classifier, the reading time for region 5 (e.g., *pencil*) in the condition with the general classifier will be faster than the reading time at the same region in the condition with a specific classifier, because building more frequent structures, of which readers have more experience, is much easier than building less frequent structures.

Third, in the postnominal constructions, the reading times at the critical region, which, as seen in Table 2, is region 5 (e.g., *three*) will be faster in the general classifier condition than in the specific classifier condition, because postnominal constructions with general classifiers occur more often than postnominal constructions with specific classifiers.

4.5. Results

This section reports the findings from the self-paced reading task. For the analyses of the task results, the dependent measures were comprehension accuracy rates and reading times at the critical regions. All data from participants whose accuracy was less than 70% were excluded. Three participants (7%) were excluded, leaving 40 participants' data to be included in the analysis. Reading times more than 2.5 standard deviations above or below the mean were removed; this affected 8.07% of the data. The analyses presented below are based on the remaining data.

4.5.1. Comprehension accuracy

The mean comprehension accuracy rate, which was calculated from the percentage of correct responses, was 86%. The average correct response percentages differed significantly in the prenominal and postnominal conditions: 83% for the prenominal constructions and 90% for the postnominal constructions ($p < .01$). This result indicates that the participants found it easier to read the postnominal constructions, regardless of the type of classifier, as predicted based on the higher frequency of postnominal structures.

4.5.2. Reading Times

The results of the reading time analysis are reported first for the prenominal condition and then for the postnominal condition. Figure 2 illustrates the differences in reading times for prenominal constructions with specific classifiers and the general classifier. At region 5 (associated nouns), the critical region for the prenominal constructions, reading times in the condition with the general classifier were significantly faster than those in the condition with a specific classifier (F_1 (1,36) = 9.641, $p < .01$; F_2 (1,23) = 4.739, $p < .05$). This result is consistent with the second prediction, which was based on the assumption that more frequent structures, such as prenominal constructions with a general classifier, are more familiar and thus easier to comprehend than less frequent structures, such as prenominal constructions with a specific classifier.
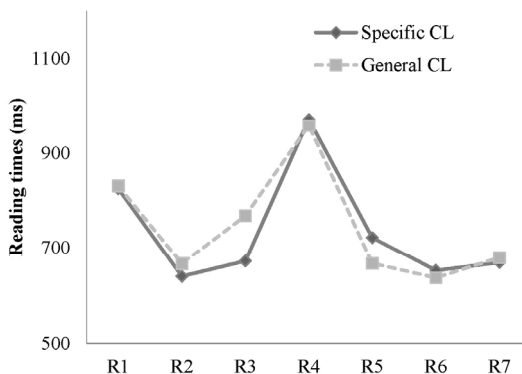

Figure 2. Reading times for prenominal constructions.

Reading times for the postnominal constructions are shown in Figure 3. At region 5 (numeral classifier), the reading times did not differ in the two classifier conditions. At region 6 (locatives), the place where any spill-over effect would be expected, reading times in the two conditions again did not differ. This result is not consistent with the third prediction. Nevertheless, it is a reasonable result because a backward retrieval process for checking agreement between classifiers and their associated nouns is difficult; that is, the classifier might not be strong enough to reactivate the information of the preceding associated nouns.
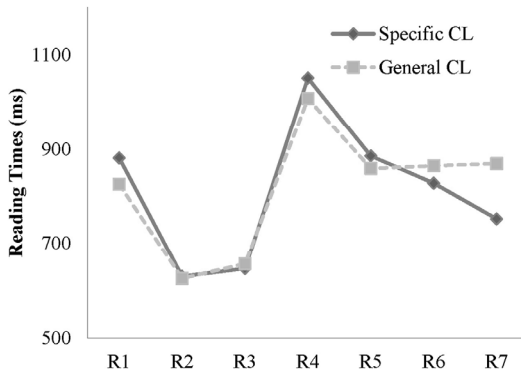

Figure 3. Reading times for postnominal constructions.

5. General Discussion

The study tested three predictions to assess the applicability of the distributional frequency-based processing account to Korean. The first prediction was borne out: Comprehension accuracy from the self-paced reading experiment showed that Korean native speakers more accurately read postnominal constructions than prenominal constructions regardless of the type of classifier (83% vs. 90%). This suggests that the readers' greater familiarity with the more-frequent postnominal constructions than the less-frequent prenominal constructions, as indicated by the corpus analysis (Figure 1), helped them to comprehend the more frequent type of sentences.

The second prediction was also borne out: In the self-paced reading task, the reading times at region 5 of the prenominal constructions in the general classifier condition were significantly faster than in the specific classifier condition. The familiarity triggered by the more frequent structures — prenominal constructions with the general classifier — facilitated the readers' comprehension of the associated nouns. These findings are similar to those of previous studies on frequency-based theories focusing on distribution (MacDonald, 1999; MacDonald & Christiansen, 2002; Reali & Christiansen, 2007a, 2007b; Gennari & MacDonald 2009; Wells et al., 2009). Comprehending rarer structures causes more difficulty in processing sentences; for example, the relative difficulty of processing object relative clauses, compared to subject relative clauses, is due to the lower frequency of object relative clauses. Assuming that experience from the input is a key factor in comprehension, more experience of particular patterns facilitates their processing. Therefore, this finding adds to the crosslinguistic evidence in support of frequency occurrence theories.

However, the third prediction was not supported by the findings. The reading times at the critical region of the postnominal constructions in the two classifier conditions did not differ, despite the greater frequency of postnominal constructions with the general classifier. One factor in this unexpected result might be the difficulty of retrieval in the online comprehension task (Lewis & Vasishth, 2005; Xiang, Polinsky, Kelly, Chen, & Wang, 2009). In postnominal constructions, the intervening words between associated nouns and classifiers (e.g., 'eraser' in Table 2) hinder the retrieval process for confirming semantic agreement between classifiers and their preceding associated nouns. A finding from Xiang et al.'s (2009) study supports this explanation: They found no evidence of a retrieval process occurring in Chinese postnominal constructions, which they ascribed to the absence of cues triggered in associated noun-classifier patterns.

In summary, the results from this study's self-paced reading task confirmed the predictions of the frequency-based approach based on distribution. The study therefore supports the account that experience of distributional frequency facilitates comprehension because building in-

frequent structures is more difficult than building frequent structures. Furthermore, this study contributes to crosslinguistic investigations of the frequency-based account by testing it in Korean.

However, three questions were raised by this study. First, there might be a factor other than frequency that creates a processing advantage for postnominal constructions. Supposing that intervening words between classifiers and nouns are structurally related to the associated nouns in prenominal constructions, the postnominal construction might be easier to process. A second puzzle concerns the finding from the second question, which suggests that the lexical frequency also might matter. In other words, not only the frequency of the structure but the frequency of lexical item *kay* also might influence the processing of the sentences. Further research is needed to tease apart these two factors. Moreover, the different frequency rates of specific classifiers, which can be observed in corpus analysis, should be addressed in further research, because different distributions of specific classifiers could lead to different reading times at the associated nouns. Finally, the third prediction still remains questionable. If readers process agreement between classifiers and nouns, there should be a difference in reading times of postnominal constructions with specific and general classifiers, but no such difference in reading times appeared. One possible explanation is that the general classifier *kay* is vague in its semantic agreement with its associated noun. Further research could examine this possibility by including ungrammatical sentences with disagreement between classifiers and their associated nouns in the materials. In spite of the remaining questions, this study's findings suggest that frequency-based processing potentially explains the preference for postnominal constructions observed in the corpus analysis — in particular, prenominal constructions with general classifiers.

6. Concluding Remarks

In this paper, I reported on a test of frequency-based processing in Korean numeral classifier constructions. An online experiment found that processing postnominal constructions is easier than processing prenominal constructions, as demonstrated by the participants' higher accuracy in their responses to comprehension questions on the sentences containing the postnominal construction. Specifically, the co-occurrence of general classifiers and prenominal constructions facilitates the processing of their associated nouns, consistent with findings in the corpus analysis. However, a question remains as to why the difference in reading times at the classifier in postnominal construction does not appear. One possible answer has to do with the difficulty of checking the agreement between classifiers and nouns in the retrieval process. Hopefully, future research will shed light on this matter.

References

Bley-Vroman, Robert and Ko, Hyunsook. (2005). *Corpus linguistics for Korean language learning and teaching*. Honolulu: University of Hawai'i Press.

Christiansen, Morten H. and MacDonald, Maryellen C. (2009). A usage-based approach to recursion in sentence processing. *Language Learning* 59, 126-161.

Clifton, Chalres Jr. and Frazier, Lyn. (1989). Comprehending sentences with long distance dependencies. In G. Carlson & M. Tanenhaus (Eds.), *Linguistic structure in language processing* (pp. 273-317). Dordrecht, Netherlands: Kluwer.

Frazier, Lyn and Fodor, Janet D. (1978). The sausage machine: A new two-stage parsing model. *Cognition* 6.4, 291-325.

Frazier, Lyn. (1979). On comprehending sentences: Syntactic parsing strategies (PhD thesis). University of Massachusetts, Amherst, MA.

Gennari, Silvia P. and MacDonald, Maryellen C. (2009). Linking production and comprehension processes: The case of relative clauses. *Cognition* 111, 1-23.

Gibson, Edward and Wu, Iris H.-H. (2013). Processing Chinese relative clause in context. *Language and Cognitive Processes* 28, 125-155.

Gibson, Edward. (1998). Linguistic complexity: Locality of syntactic dependencies.

Cognition 68.1, 1-76.

Gibson, Edward. (2000). Dependency locality theory: A distance-based theory of linguistic complexity. In A. Marantz, Y. Miyashita, & W. O'Neil (Eds.), *Image, language, brain* (pp. 95-126). Cambridge, MA: MIT Press.

Grodner, Daniel and Gibson, Edward. (2005). Consequences of the serial nature of linguistic input for sentential complexity. *Cognitive Science* 29, 261-291.

Hale, John. (2001). A probabilistic Earley parser as a psycholinguistic model. In *Proceedings of the 2nd Meeting of the North American Chapter of the Association for Computational Linguistics on Language Technologies* (pp. 159-166). Pittsburgh, PA: Association for Computational Linguistics.

Hsiao, Franny and Gibson, Edward (2003). Processing relative clauses in Chinese. *Cognition* 90, 3-27.

Jäger, Lena, Chen, Zhong, Li, Qiang, Lin, Chien-Jer C., and Vasishth, Sjravan. (2015). The subject-relative advantage in Chinese: Evidence for expectation-based processing. *Journal of Memory and Language* 80, 97-120.

Just, Marcel A. and Carpenter, Patricia A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review* 99.1, 122-149.

Just, Marcel A., Carpenter, Patricia A., and Woolley, Jacqueline D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General* 111.2, 228-238.

Kim, Jong-Bok and Yang, Jaehyung. (2006). Processing Korean number classifier constructions in a typed feature structure grammar. Ms. Kyung Hee University, Kangnam University, Korea.

Ko, Heejeong and Oh, Eunjeong. (2012). A hybrid approach to floating quantifiers: some experimental evidence. *Linguistic Research* 29.1, 1-40.

Ko, Heejeong. (2007). Asymmetries in Scrambling and Cyclic Linearization. *Linguistic Inquiry* 38, 49-83.

Lee, On-Soon. (2014). The frequency effect on L2 English learners' sentence comprehension: The case of objective relative clauses. *Modern English Education* 15.3, 123-139.

Levy, Roger and Keller, Frank. (2013). Expectation and locality effects in German verb-final structures. *Journal of Memory and Language* 68, 199-222.

Levy, Roger, Fedorenko, Evelina, and Gibson, Edward. (2013). The syntactic complexity of Russian relative clauses. *Journal of Memory and Language* 69, 461-495.

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

Lewis, Richard L. and Vasishth, Shravan. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cognitive Science* 29, 1-45.

Lewis, Richard L., Vasishth, Shravan, and Van Dyke, Julie. (2006). Computational principles of working memory in sentence comprehension. *Trends in Cognitive Sciences* 10.10, 447-454.

Linzen, Tal and Jaeger, Florian T. (in press). Uncertainty and expectation in sentence processing: Evidence from subcategorization distributions. *Cognitive Science*.

MacDonald and Maryellen, C. (1999). Distributional information in language comprehension, production, and acquisition: three puzzles and a moral. In B. MacWhinney (Ed.), *The emergence of language*. Mahwah, NJ: Erlbaum.

MacDonald, Maryellen, C., and Christiansen, Morten H. (2002). Reassessing working memory: comment on Just and Carpenter (1992) and Waters and Caplan (1999). *Psychological Review* 109, 35-54.

Park, Myung-Kwan, Chung, Wonil, and Choi, Kiyong. (2014). kwuke yutong swu-pwunlyusa (floating numeral classifier)uy silhemenehakcek cungke: ERP kipanyenkwu. *Studies in Generative Grammar* 24, 231-254.

Reali, Florencia and Christiansen, Morten H. (2007a). Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language* 57, 1-23.

Reali, Florencia and Christiansen, Morten. H. (2007b). Word chunk frequencies affect the processing of pronominal object-relative clauses. *The Quarterly Journal of Experimental Psychology* 60, 161-170.

Sohn, Ho-min. (1999). *The Korean language*. UK: Cambridge University Press.

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

Unterbeck, Barbara. (1994). Korean classifiers. In Young-Key Kim Renard, ed., *Theoretical Issues in Korean Linguistics*, 367-386. Stanford: CSLI Publications.

Wells, Justine. B., Christiansen, Morten. H., Race, David. S., Acheson, Daniel. J., and MacDonald, Maryellen. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology* 58, 250-271.

Wu, Fuyun, Kaiser, Elsi, and Andersen, Elaine. (2009). The effect of classifiers in predicting Chinese relative clauses. In Michael Grosvald & Dionne Soares, eds. *Proceedings of the Western Conference on Linguistics* (WECOL), 318-329. Davis, University of California.

Xiang, Ming, Polinsky, Maria, Kelly, Christina, Chen, Lan, and Wang, Suiping. (2009). The effect of partial semantic feature match in forward prediction and backward retrieval. *In poster presented at the 22th Annual CUNY Sentence Processing*

Conference. UC Davis, CA.

Yun, Hongoak and Hong, Upyong. (2014). The effect of role predictability and word predictability on sentence comprehension. *Journal of Cognitive Science* 15, 345-390.

Yun, Hongoak, Nam, Yunju, Yoo, Duck Geun, and Hong, Upyong. (2015). The effect of role shifting and expectation in the processing of center-embedded relative clauses in Korean. *Linguistic Research* 32.2, 313-353.

On-Soon Lee
Department of English Language and Literature
Korea University
145 Anam-ro, SeongBuk-gu
Seoul 136-701, Korea
Email: onsoon@gmail.com

Appendix A

	Classifier Type	Prenominal Constructions	Postnominal Constructions
Specific	kok (song)	0	2
	phyen (film)	0	23
	calwu (pencil)	0	11
	tay (car)	0	14
	pyeng (beer)	0	15
	kwen (book)	0	2
	cang (paper)	0	9
	pel (shirt)	0	11
	can (coffee)	0	23
	kulwu (tree)	0	0
	songi (flower)	0	34
	thong (bucket)	0	0
General	kay (inanimate entities)	122	308