Effects of Visual, Lexical, and Contextual Factors on Word Recognition in Reading Korean Sentences

Say Young Kim[1, 2] and Donald J. Bolger[3]

[1]*National University of Singapore, Singapore*
[2]*Sejong University, Korea*
[3]*University of Maryland, College Park, USA*
sayyoung.kim@gmail.com, djbolger@umd.edu

In order to assess the role of visual, lexical, and contextual information on word identification during Korean sentence reading, a self-paced reading experiment was conducted. It was found, with regard to word length variables, that the number of syllables and the number of visual features affected reading times significantly, but the other sub-lexical units (i.e., phonemes and letters) did not. The findings suggest that when taking internal structure variations into account, the relevant processing unit in Korean in the context of sentence is the syllable. In addition, the main effects of both word frequency and predictability on reading time were significant, respectively; however, the interaction between these two variables was not. The results imply that Korean word recognition during sentence reading is affected by word frequency and word predictability, additively.

Key words: *word recognition; word length; word frequency; predictability, Korean*

1. Introduction

The written language system of Korean is unique from other western alphabetic writing systems in several ways. As we will discuss, the orthographic system of Korean, Hangul, codes for sound at both the level of the phoneme and the syllable simultaneously. Moreover, the Korean

Journal of Cognitive Science **18-1: 43-83, 2017**
Date submitted: 03/06/17 Date reviewed: 03/29/17
Date confirmed for publication: 04/03/17

Hangul system has more directly embedded the morphology of the spoken language. That is, the spatial packaging of the written syllables, the kulja, reflect more than just the phonological structure (e.g., 사 "sa"), but will preserve the morphological variance of that same syllable (e.g., 운전-사 "wuncen-sa" meaning *drive-er*). Lastly, the syntactic structure of the spoken/written language is more flexible than systems like English that have definitive word order. Thus, the influences on visual word processing in Korean may be fairly unique from systems like English particularly in the context of processing written discourse. The aims of the present study are to examine whether visual word processing in Korean operates under similar orthographic, lexical, and contextual constraints as other written language systems like English or do the variations in writing systems result in unique patterns for Korean. Moreover, the present study focuses on word processing during sentence reading, which is more ecologically valid of naturalistic reading than processing words in isolation which tend to reflect meta-linguistic processes that may be more sensitive to task-specific demands (Balota & Yap, 2006).

In order to address these aims, we first explore how the nature of a language and its writing system differentially affect the word identification process (Perfetti, 2003; Taylor, 1980; Frost, Katz, & Bentin, 1987). For example, although word length is consistently shown to be a primary predictor in word identification in English, this effect may depend on the nature of the orthographic structure of the language. As stated earlier, the Korean writing system has both alphabetic and syllabic properties unlike English, which is alphabetic but not syllabic. However, little research has been conducted to determine how factors such as word length, word frequency, syllabic and subsyllabic features affect the processing of Korean words in sentences. For instance, the letter or sub-syllabic features such as the body-coda (CV-C) have been shown to provide additional constraints on the recognition and processing of Korean words in isolation (Kim & Bolger, 2016; Yoon, Bolger, Kwon, & Perfetti, 2002). Given the efficiency afforded by this transparent alphabetic system, it is possible that such factors affect processing written words on-line in the context of larger discourse.

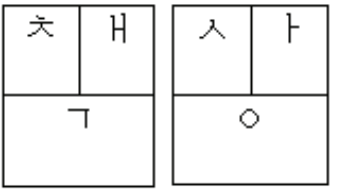
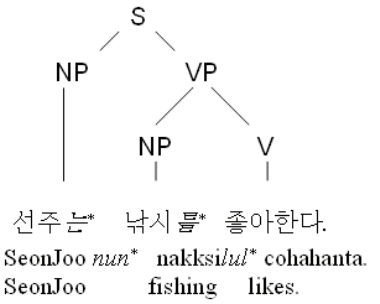


Figure 1. Spatial arrangement of letters in Korean syllables and words. The example in the figure shows the two-syllable word "chayk sang" (desk in English) with the left-most "box" containing the first CVC syllable (i.e., the letters, ㅊ, ㅐ, ㄱ), and the right-most "box" containing the second CVC syllable (i.e., the letters, ㅅ, ㅏ, ㅇ). The boxes are shown to illustrate the spatial arrangement of the individual letters in the syllables.

1.1 The Korean writing system and sentence structure

The Korean alphabet, Hangul, consists of 24 primary letters, with 14 consonants and 10 vowels. It is considered an alphabetic syllabary because each phoneme is represented by a letter, and each word is made from letters that are combined into syllable packages that consist of compact character blocks as in Figure 1 (Taylor, 1980). Thus, syllable boundaries in Korean words are visually salient. Morpho-syntactically, Korean is considered to be a syllabic system that is largely composed of polysyllabic morphemes with an elaborate inflexion system that is based on suffixes.

The example in Figure 1 divides the word "책상" (pronounced /*chayk sang*/, meaning *desk* in English) into its parts. Each syllable is also written in a square. As the figure shows, each syllable begins with a consonant, either pronounced or silent (i.e. the character, X, is used as an empty set place holder for open syllables), and has at least one additional consonant and one vowel, which may represent a CV, CVC, or CVCC syllable. Because of its transparent, one-to-one correspondence between graphemes and phonemes, Korean generally lacks the inconsistent mappings that exist in the English alphabetic system. For instance, in English, a single letter can represent a variety of sounds, and a single sound can be represented by several different letters or letter group (Taylor & Taylor, 1983). The syllable package in Korean orthography, the *kulja*, can be decomposed into a number of strokes like Chinese. In Chinese, the stroke within a



Note: * indicates the case markers for subject and object, respectively.

Figure 2. Example sentence showing the head final structure of Korean.

character indeed increases the complexity that may elicit differential level of processing load (Yeh & Liu, 1972). Thus, it is questionable whether the numbers of strokes (a measure of orthographic complexity) affect word recognition in Korean.

At a higher level, Korean sentence structure is different from many other commonly studied languages (e.g., English) in that it is head-final, which simply means that verbs are placed at the end of sentences (i.e., SOV word order; see Figure 2). Therefore, an extensive set of noun phrases (e.g. subject, direct object, indirect object) and modifiers occur prior to the verb in a sentence. Moreover, the use of case markers and lack of articles allow for word order in Korean to be more flexible than in English, and syntactic categories are frequently labeled by case markers (e.g., '는' /-nun/ for a subject and '를' /-lul/ for an object in Figure 2). These important characteristics of Korean may influence the word predictability effect because it is determined, at least in part, by the syntactic information that occurs prior to a given word. Thus, one might predict that, in the case of Korean, the head-final syntactic structure and the use of case markers might together provide relatively weak contextual constraints on each word in a sentence, especially weak syntactic constraints, and thereby reduce the overall size of any word predictability effects that might be observed in the present experiment.

1.2 The effect of word length

Word length is thought of as a fundamental constraint to identify a word (New, Ferrand, Pallier, & Brysbaert, 2006 for review), and its universal importance has been suggested by results of many different languages, such as Korean (Park, 1993; Lee & Kim, 1989; Nam, Seo, Choi et al., 1997), English (Lee, 1999; Lima & Pollatsek, 1983; Spoehr & Smith, 1973; Prinzmetal, Treiman, & Rho, 1986), Hebrew (Lavidor & Whitney, 2005), and French (Juphard, Carbonnel, & Valdois, 2004). These studies have generally found longer latencies for naming and lexical decision for target words that are longer in length both orthographically and phonologically. This result is thought to be due to the fact that our visual system has limited visual acuity which makes it more difficult to identify larger objects (e.g., words). In addition, the tasks that have been used (e.g., naming) contain particular metalinguistic demands, for instance, matching the printed scripts to their sounds. If the time to complete these processes increases linearly as the number of characters increases (which would be true as word length increases), then it would take more time to name long words than short words. In relation to word length, one might ask whether or not other variations within a word affect reading process. One possible variation within a word can be its internal structure, that is, density or complexity. Different ways of writing words might result in different levels of visual complexity.

In relation to the processing unit during word recognition, some studies using Korean found the role of word length, for instance, to be an effect of the number of letters (Park, 1996) or the number of letters and strokes (Nam et al., 1997) using word in isolation tasks (e.g. naming or lexical decision). However, whether these processes will persist with different types of tasks (e.g., silent reading of words in sentences) is still unknown. Reading a word in a sentence might affect the role of sublexical units in single-word recognition. In addition, the orthographic properties of Korean, which is printed not as a linear arrangement, but as a packaged arrangement within a character space, might affect this process in a sentence reading situation. For example, when native-Korean speakers read a sentence, they might be less affected by detailed features within a single word possibly because details in a square space of syllable can be processed in parallel.

Table 1. Example of Hangul Syllable Blocks Showing Three Levels of Complexity (adapted from Taylor, 1980) and Number of Visual Features

Complexity Level	Linear Arrangement	Syllable-Block	Number of visual feature	Syllable and pronunciation	Meaning
I	ㅏ	아	3	V /a/	suffix; ah
I	ㄷㅏ	다	5	CV /da/	all
II	ㅏㄹ	알	8	VC /al/	egg
II	ㄷ ㅏㄹ	달	10	CVC /dal/	moon
III	ㄷ ㅏㄹㄱ	닭	12	CVCC /dalg/	hen

In terms of the internal structure of Korean words, there are two properties of the Korean alphabetic (*Hangul*) system: the first is the different types of CV blocks that can be used to construct words (Taylor, 1980); the second is the number of visual features making up letters and words. In Hangul, the square-like spatial arrangements of the individual letters within each syllable may possibly relate to the complexity of a word, because there can be relative variation of the visual features or the number of CV block types in the limited space. For instance, words consisting of the same number of syllables can consist of different CV blocks, and thus contain different numbers of visual features. According to Taylor (1980), there are five different CV combination types that can be categorized into three levels of visual complexity (see Table 1).

Basically, the complexity increases by adding a consonant. These levels of complexity might be predicted to affect word recognition simply because increasing the complexity of a word increases the number of visual features, and thereby induces more low-level visual processing necessary to identify the word. Therefore, it may be worthwhile to test the role of complexity during Korean word recognition.

With respect to the internal structure of Korean words, visual features should be considered as a component of processing units. The visual features of Hangul are similar to the stroke patterns in Chinese (Wang, 1981). Nam and colleagues (1997) examined the role of sublexical units in Korean as well as the visual features, such as horizontal, vertical, and

diagonal lines, and circles (see Table 1). Their naming and lexical-decision experiments showed that the reaction times in these tasks increased as the number of visual features in the words increased, even when the number of syllables was controlled. These results may reflect the fact that native speakers of Korean are sensitive to the visual complexity of words, even though these visual features are not related to the linguistic aspects of the language.

Based on the internal structure of a Korean word, one might still ask whether or not the CV blocks and/or visual components of the letters (e.g., vertical lines) have a role in word recognition within text. Although two previous studies dealt with characteristics of the internal structure of Korean, neither tested its effect on word recognition during sentence reading. Taylor (1980) used a non-language task and Nam et al. (1997) used recognition task of isolated words. Task differences might be expected to induce a specific priority in word recognition, for example, compared to silent reading task, the naming task possibly requires that readers access phonological information from printed words.

1.3 The effect of word frequency

The second variable considered in the present study is frequency of occurrence. Many previous studies using a variety of different experimental methods have consistently demonstrated word frequency effects (Balota & Chumbley, 1984; Monsell, Doyle, & Haggard, 1989; Allen, McNeal, & Kvak, 1992). Allen et al. (1992) examined the effect of word frequency by manipulating stimulus onset asynchrony (SOA), the interval between onset of the prime and the onset of the target. In their lexical decision experiments, word-frequency effects were found across the different SOAs, between the presentation of a letter string and a subsequent pattern mask in a lexical decision task, with no interaction between frequency and SOA. In addition, Monsell et al. (1989) designed a series of experiments to compare the effect of word frequency using different types of tasks (e.g., semantic categorization, syntactic categorization, etc.). The results showed a consistent word-frequency effect in lexical recognition with little variability of its effect size. Together, these results suggest that word frequency effects are robust and common.

A more interesting aspect of frequency could be its relationship with other variables in a word. For example, Lee (1999) found a main effect of word frequency and an interaction between frequency and word length with native English speakers, with larger word-length effects for low-frequency than high-frequency words. Studies of Korean have also reported this kind of interaction (Park, 1993; Nam et al., 1997). For example, Park (1993) showed a larger frequency effect in 2- and 3-syllable words than in 1- and 4-syllable words. The interaction between word frequency and word length suggests the locus of both variables (Balota & Chumbley, 1985). Namely, word length affects not only prelexical processes, but also lexical processes, because the interaction between these two variables can be interpreted as evidence that both variables share a processing stage (Sternberg, 1969).

1.4 The effect of predictability

Finally, the contextual effect on word recognition is examined in the present study (i.e., word predictability). Word predictability has been studied using different methods like eye tracking (Rayner & Well, 1996) which examined the effect of contextual constraint on the eye movements of readers who encountered predictable versus unpredictable words in a sentence. Word predictability was defined as the amount of contextual constraint which was manipulated by cloze probability. In order to determine predictability, cloze probability was assessed in a separate group of participants that were asked to complete sentence fragments by filling in a target word. Low-constraint (i.e., unpredictable) target words induced longer gaze durations (the sum of all fixations on a word prior to movement to another word) as compared to medium- or high-constraint (i.e., predictable) target words. Participants were also more likely to fixate (i.e., have greater fixation probability) on words in the low-constraint than medium- or high-constraint conditions.

Another point about predictability that is relevant to the present study is that its effect varies considerably with stimulus quality; that is, the more degraded the stimulus, the greater the effect of context (Stanovich & West, 1983). A recent study using Korean also reported a similar interaction between predictability and visual degradation. Lee (2004) manipulated the predictability of target words in sentences and asked native Korean speakers to make lexical decisions on target words that were visually

degraded or normal. The results showed that performance improved when target words were highly predictable or visually normal, with the context effects being larger when the target words were visually degraded than when they were normal. This interaction leads to a prediction that predictability affects early stages of visual word recognition, namely, a prelexical processing stage. However, predictability is also expected to influence later stage of lexical processing by confirming the word meaning based on contextual information. Several eye-tracking studies in English (Rayner, Ashby, Pollatsek, & Reichle, 2004; Huck, Thompson, Cruice, & Marshall, 2017) or French (Miellet, Sparrow, & Sereno, 2007) showed that predictability does not interact with frequency, which suggests that these two variables influence different stages of word recognition, additively. This result is interesting because it is in contrast to the prediction that the effect of contextual information can be expected to be greater on less frequent words.

In sum, previous studies relevant to three main variables (i.e., word length, word frequency, and word predictability) have been reviewed. Because of the Korean writing system and orthography, internal structure properties should be considered (i.e., visual features and complexity). Together, these word length and internal structure properties will be investigated to examine what levels of sublexical units in Korean play a significant role during reading. In addition, word frequency and word predictability effects will also be investigated in the following two experiments.

The present study examines how processing written words in Korean is affected by three main variables: word length, frequency, and predictability. To understand whether the Korean writing system is susceptible to these factors of word processing, we employed a self-paced reading paradigm addressing the following four research questions: 1) What are the salient units of orthographic processing (e.g., letters or syllables) in the processing of Korean words and do they play a role in online visual word processing? 2) To what degree do lexical variables (e.g., word frequency) play a role in the reading of Korean words in text? Third, how much do the sentential constraints imposed by syntax and/or meaning (i.e., word predictability) affect word processing during the reading of Korean? Finally, how do the three main variables of interest (word length, frequency, and predictability)

interact with each other?

2. Method

A cloze task was conducted to determine predictability norms for the stimuli. In the self-paced reading task, participants were asked to read sentences containing target words that were selected so as to vary along three main dimensions: word length, word frequency, and word predictability.

2.1 Cloze task

The objective of this cloze task was to select a set of target words that were highly predictable from their sentence contexts so that these items would be used in a self-paced reading experiment. Participants were presented sentence fragments that terminated with a blank (See Example Sentences). Participants were instructed to write down the first word that come to their mind after reading each fragment.

> Example Sentence
> Sentence fragment: 용욱이는/ 밤새워/ 시험 공부를 하기/ 전에/ 한잔의 _____.
> Pronounced: Yongwuk-i-nun/ pam-say-we/ si-hem kong-bu-lul ha-ki/ cen-ey/ han can-uy
> Phrasal meanings: Yong-Wook/ during overnight/ studying for exam/ before/ a cup of _____.

As shown in the example sentence, the main verb of the sentence was left blank as the verb in a Korean sentence, which is head-final structure, is frequently placed at the end of the sentence (after the target word) and thereby it might be thought to provide little constraint on the target word. Due to this head-final structure of Korean, the stimuli for this cloze task were designed to contain enough semantic constraint before the blanks (i.e., target words' positions) to explicitly manipulate the degree of predictability. In contrast, the verb in an English sentence is typically located near middle of the sentence, and thereby provides relatively more constraint on the

target word.

2.2 Participants

Seventy-six native Korean speakers were recruited from a major university in Seoul, Korea. They were paid for their participation.

2.3 Stimuli and procedure

One hundred twenty partial sentences were presented on sheets of paper with each sentence fragment containing a blank space. This task was conducted as a pencil-and-paper test. Participants were asked to fill in the blank with the first word that comes to mind. The duration of the task was about 20 minutes.

3. Results

Thirty-two responses from the cloze task that were given more than 25% of the time in sentence fragments were assigned to the *high-predictability* condition (M = 60%; range = 25-97%). Note that the absolute range of words that were judged to be "predictable" in Korean was less than in the results of previous studies using English (Rayner et al., 2004). The mean cloze-probability for highly predictable target words in Korean was 60%, whereas in English it was 78% (Rayner et al., 2004). One possible reason for this difference is that the head final structure of Korean sentences promotes a more flexible word order than in English (e.g., Inoue, 1991; Mazuka & Lust, 1990) and may thereby reduce the effect of contextual constraint that comes from syntactic and/or semantic information. In other words, English sentences can provide both semantic and syntactic constraints on a target word in a sentence, but Korean sentences mostly provide semantic, not syntactic constraints on a target word.

3.1 Self-paced reading experiment

To investigate how word length, frequency, and predictability affect word reading in Korean, it is first helpful to explain how the properties of the Korean writing system and sentence structure would influence natural reading. First of all, a word-length effect is expected, as has been found

in previous studies (Lee & Kim, 1989; Nam et al., 1997). However, the effective processing units (e.g., letters, phonemes, and/or syllables) may be different from previous research using Korean because the task involves sentence reading and not naming or lexical decisions of isolated words. This prediction is plausible, if we can assume the following: When people read a sentence or a text silently, they might not need to activate phonological information of every single word explicitly (compared to a naming task), nor do they need to judge whether every given letter string is a word or not (compared to a lexical decision task). In addition, the internal structure of the Korean word, the type of CV blocks and the number of visual features, can affect the latency for word recognition, because of the packaged written form of Korean. Sometimes, different CV types (adding C or V) lead to additional phonological information, but not all the time. For instance, the last consonant "C" in CVCC usually only differentiates the words orthographically, not phonetically. Therefore, this complexity level can be tested to see Korean native speakers' sensitivity about the Korean syllable's internal structure. If native Korean speakers are sensitive to the internal structure of Korean, the complexity level could affect the reading time of words. Nam and colleagues (1997) demonstrated that latencies for naming and lexical decision for Korean words increased as the number of visual features increased.

The second prediction is that an effect of word frequency is also expected because it has been consistently shown across a number of experiments and tasks. This prediction is based on the traditional definition of word recognition as a pattern recognition process in which the visual input is encoded and looked up in the mental lexicon to find the best match (Tsai, 2001).

Finally, an effect of word predictability is predicted even if Korean has a more flexible word order than English, which means that the words preceding the target words could be thought to provide syntactically weaker constraints compared to English.

As well as three main effects, we also predict there will be interactions among the variables. Previous studies of Korean (Park, 1993; Nam et al., 1997) showed an interaction between word length variables (letters or syllables) and word frequency. In particular, Park (1993) reported an

interesting pattern of interaction that depended on the task. Park used one to four syllable words with high- and low-frequency words. In the naming task, reaction time increased as the number of syllables increased in 1–4 syllable words, but the frequency effect was larger in two and three syllable words (21 ms vs. 31 ms) than in one-and four-syllable words (1 ms vs. 10 ms). This interaction between the number of syllables and word frequency was significant. A similar pattern of results was found for the lexical decision task – both the main effects of syllable and of word frequency, and the interaction were significant, but the reaction time for one-syllable words was longer than two-syllable words. Park (1993) suggested that a possible reason for this pattern of reaction times might be semantic uncertainty of one-syllable words in Korean. This is because one-syllable words may have many homophones, and most syllable units (kulja) are combined into multiple-syllable words. Therefore, it is unclear whether this result will replicate in this present experiment using words in a sentence. If the sentential context is reliable, then the semantic uncertainty of one-syllable words can be reduced, and the reading time may be faster than for two-syllable words. In his results, the interaction pattern in the lexical decision task is consistent with the result of the naming task, and it is more salient. Frequency effects were greater in two-and three-syllable words (82 ms vs. 62 ms) than in one-and four-syllable words (41 ms vs. 30 ms). Based on the previous studies, we predict that an interaction between number of syllables and word frequency will be found in the current experiment. However, the interaction pattern may be modulated by the effect of context in this current experiment.

For the purpose of examining how the visual, lexical, and contextual characteristics of Korean affect word recognition during natural reading, word length, frequency, and predictability were manipulated among the target words. To do this, three different sets of stimuli were constructed as described in the next section.

3.2 Participants

Twenty-one native speakers of Korean in the Pittsburgh community participated in this experiment, and twenty-five Koreans in Seoul also participated. All of the participants had normal or corrected-to-normal

vision and received monetary compensation for their participation.

3.3 Stimuli

A set of 108 sentences was generated, and part of this sentence set originated from the cloze-task (see stimulus Set III). These stimuli consisted of three sets of 32 sentences and 12 filler sentences. Only a small number of fillers were included because each set was expected to function as fillers for the other sets. More specific information about each set of sentences is given in the following:

Set I consisted of 32 sentences containing target words that varied in terms of word length (i.e., number of syllables: 1-4) and word frequency (high frequency: *Mdn* = 180 per million,[1] range = 59-849; low frequency: *Mdn* = 1.3 per million, range = 1-26). Therefore, the properties of the target words in Set I provided the opportunity to evaluate the effect of the number of syllables as a possible unit of word length, and a possible interaction with frequency. Although there are other ways to define word length (e.g., number of letters), the number of syllables in Korean is strongly correlated to these other word-length measures (Nam et al., 1997). Therefore, only the number of syllables was considered in this first stimulus set; the possible roles of the other units of word length are examined in Set II, while the number of syllables is controlled.

Set II consisted of 32 sentences containing two-syllable target words with various numbers of phonemes (*M* = 5.1; range = 3-7) and letters (*M* = 2.6; range = 2-6). The internal structure properties of the Korean words also varied in this set, but they were not directly manipulated (as was frequency in Set III) based on two properties: the various possible syllable-block (CV) combinations (Taylor, 1980), and the number of visual features (number of strokes). As Table 1 shows, CV block combinations can have three levels of complexity based on a single letter, so the complexity of the present stimuli was based on the sum of each syllable's complexity level (*M* = 4.8; range = 3-6). The number of visual features was also based on the sum of all components in each word (*M* = 12.7; range = 7-19).

---

[1] The database referred to here provide the frequency based on 1.5 million, but it is converted based on 1 million for comparing with other general data corpora.

Table 2. Examples of Target Stimuli for the Cloze Task

Condition	Example Sentences
HF–P	일요일에는 미뤘던 빨래와 **청소**를 해야겠다. On Sunday, postponed, laundry, and, **cleaning**, have to.
HF–U	차가 더러워졌으니 일요일에는 반드시 **청소**를 해야겠다. Car, dirty, on Sunday, necessarily, **cleaning**, have to.
LF–P	차가 더러워졌으니 일요일에는 반드시 **세차**를 해야겠다. Car, dirty, on Sunday, necessarily, **car wash**, have to.
LF–U	일요일에는 미뤘던 빨래와 **세차**를 해야겠다. On Sunday, postponed, laundry, and, **car wash**, have to.

Notes: Bold indicates target words; HF: high-frequency targets; LF: low-frequency targets; P: predictable targets; U: unpredictable targets.

Set III consisted of 32 target sentences extracted from the cloze task and containing four types of target words: (1) high-frequency predictable words; (2) high-frequency unpredictable words; (3) low-frequency predictable words; and (4) low-frequency unpredictable words. The frequency of the high-frequency words was over 53 per million (M = 149, Mdn = 114, SD = 134) and the frequency of the low-frequency words was less than 9 per million (M = 2.9, Mdn = 2.3, SD = 2.5) based on Korean frequency database (National Academy of the Korean Language, 2002). Predictable target words were selected from the responses in the cloze task that were given more than 25% of the time (M = 60.38%, Mdn = 63%, SD = 17.64%).

In sum, as Table 2 shows, each target word (high- or low- frequency) was embedded in sentences that made the words either predictable or unpredictable. In the unpredictable conditions, target words were selected from responses that were given less than 1% of the time. To control word length, all of the target words in Set III were two-syllable words (except for two of the items, which were three-syllable words).

Fillers consisted of 12 sentences that were presented as longer, multi-word units (e.g., phrase by phrase or a whole sentence at a time) as compared to the experimental sentences to prevent participants' automatic pressing of the button without their possessing or comprehension of the individual words.

3.4 Procedure

Participants completed a questionnaire about their gender, age, college major, and language background, and then they were instructed to begin the self-paced reading task. The sentences were presented one at a time on a computer program designed for self-paced reading experiments.[2] The initial display consisted of a line of dashes preserving spaces between words. The participants were instructed to press the space bar to see the first word, and then to press the bar to see each new word of the sentence; thus, whenever participants pressed the space bar, each set of dashes changed to a word (i.e., a self-paced moving window paradigm where each new word replaced a set of dashes). During this process, a presented word is changed to a set of dashes when participants pressed the space bar to see the next word. When only one word at a time is presented, this procedure yields a processing time measure for each word in the sentence. Participants were also instructed to read at a natural rate and to comprehend what they were reading. Before starting the experimental session, each participant read five practice sentences to become familiar with the procedure. During the experimental session, true/false comprehension questions were randomly presented after one fourth of the sentences (on average) to impose on-line comprehension of the sentence and to avoid "mechanical" pressing of the button to move forward through the text. They answered these questions by pressing the "F" key for "yes" or the "J" key for "no". They were then given feedback if their answer was incorrect. The computer automatically displayed the onset of the next sentence following each sentence or comprehension question.

4. Results

All reading times for target words that were more than 3 *SD*'s from the mean of each stimulus set were excluded from the data analyses. Two participants from the Seoul group were also excluded from the data analyses because they had already participated in our cloze task, and three participants (two from Pittsburgh and one from Seoul) were excluded

[2] The program was written by Douglas Rohde and can be obtained from: http://tedlab.mit.edu/~dr/Linger

because they had a significant number of long reading times (i.e., 33% more than 3 *SD*'s from the mean). Therefore, 41 participants' data were analyzed. All of the reading times for target words are re-calculated by dividing the mean times by the number of syllables in each target word. (This was necessary because Korean words in a sentence are typically combined with case markers; e.g., *un, nun, i, ga* for subjects, and *ul, lul* for objects. Details of this procedure are reported with the results of each stimulus set.)

4.1 Set I: Word length (number of syllables) and frequency

Before analyzing the data of Set I, 1.3% of the data more than 3 SD's from the mean were excluded. Unfortunately, there were not equal numbers of 1-4 syllable words in this stimulus set. Therefore, a hierarchical regression analysis was conducted using the order as the number of syllables, corrected frequency (i.e., logarithmic frequency), and the interaction between those two variables as predictors. The results are shown in Table 3. The R^2 statistics showed that the number of syllable did significantly predict reading times [$R^2 = .832$, $F(1, 30) = 148.\ 05$, $p < .001$].

As predicted, the effect of word length was reliable, with the reading times for target words increasing with the number of syllables. This result demonstrates the importance of word length in word recognition, and the significance of the syllable as a critical processing unit in Korean. Word frequency also significantly accounted for variance [$\Delta R^2 = .026$, $F(1, 29) = 5.20$, $p < .05$]. Finally, interaction between the number of syllable and frequency also accounted for significant proportion of variance [$\Delta R^2 = .021$, $F(1, 28) = 4.93$, $p < .05$]. Figure 3 shows the result of reading times dependent on word frequency and number of syllables. In general, as far as word frequency is concerned, high-frequency words received shorter

Table 3. Multiple Regression Analysis Results from Set I

Variable Input Order	R	R^2	Changes in R^2	F
1. Number of Syllable	.912	.832	.832	148.05***
2. Frequency	.926	.857	.026	5.20*
3. Number of Syllable × Frequency	.936	.879	.021	4.93*

Notes: * refer to $p < .05$, ***, refer to $p < .001$

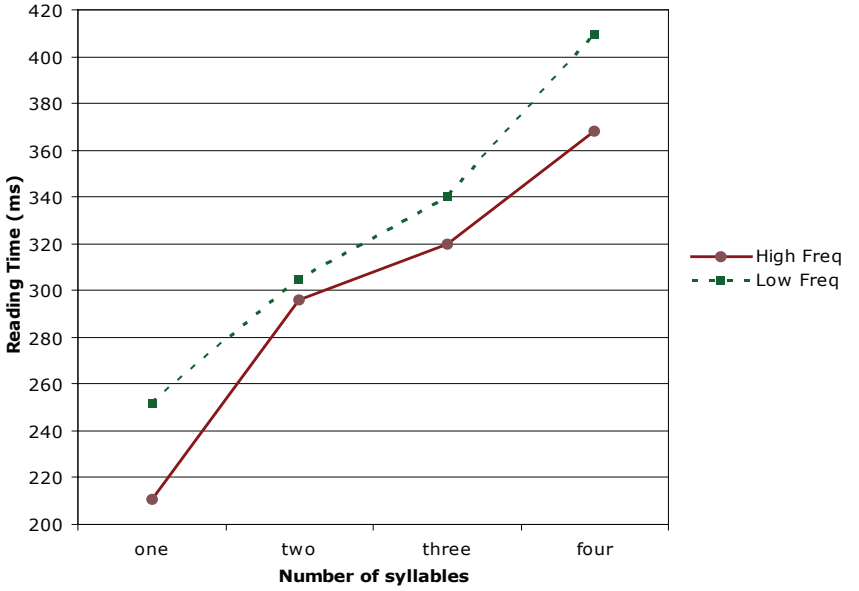


Figure 3. Reading times for target words in Set I as a function of the word frequency and the number of syllables. Reading times linearly increase as the number of syllables increases, and the latencies are faster for high-frequency words than for low-frequency words. An interaction indicates that the frequency effects are larger for one-and four-syllable words than two- and three-syllable words.

reading times than low-frequency words (M = 298.9 ms vs. 326.9 ms). This result replicates findings that frequent words were represented more accessibly status than infrequent words, as many previous studies have shown.

In addition, the mean differences between high- and low- frequency words across syllables imply that the frequency effect size is different depending on the number of syllables. This result is basically consistent with interactions between frequency and length as shown in previous English and Korean studies using naming and lexical decision (Lee, 1999; Park, 1993; Nam et al., 1997). Interestingly, however, the pattern of the effect sizes of frequency in this experiment is different from a previous study (Park, 1993) which also reported the interaction between those two variables

in Korean. Park (1993) found effects of word length (defined as the number of syllables), word frequency, and an interaction between these two factors when participants were asked to name and make lexical decisions about target words. However, there is one difference between Park's (1993) results and those of the present experiment: Park (1993) showed a larger frequency effect for two- and three-syllable words compared to one- and four- syllable words, but he did not suggest any interpretation about this tendency of frequency effect.

In contrast, the present experiment showed the different pattern of interaction as Figure 3, namely that word frequency effects were greater in one- and four- syllable words than in two- and three- syllable words. For the different effect sizes of frequency depending on the number of syllables, we examined the numerical ratio of 1–4 syllable words in the database that was used in this study. The ratio is 1: 17: 14: 5, respectively. Therefore, we can conclude that the larger effect of frequency in one and four syllable words in the present experiment is because one and four syllable words are expected to be used relatively less frequently as compared to two-and three-syllable words, and it possibly made Korean native speakers' sensitivity to one-and four-syllable words much higher compared to two-and three-syllable words. It is possible, though, that latency of word recognition is a function of not only the syllabic types (1-4 syllable), but also token frequency of each syllabic type.

4.2 Set II: Word length (number of letters, phonemes) and visual complexity

1.5% of data more than 3 *SD* from the mean were excluded from the analyses. Because each of the target words of Set II was a two-syllable word, only the number of phonemes (and letters, since in Korean the two are highly correlated) varied across the target words (in contrast to the target words in Set I). The visual complexity level of each word was also defined using both the CV block measure (see Table 1) and the number of visual features (Nam et al., 1997).

A hierarchical regression analysis as an order of the number of letters, phonemes, and visual features, along with visual complexity showed that only the number of visual features reliably predicted reading times (*r*

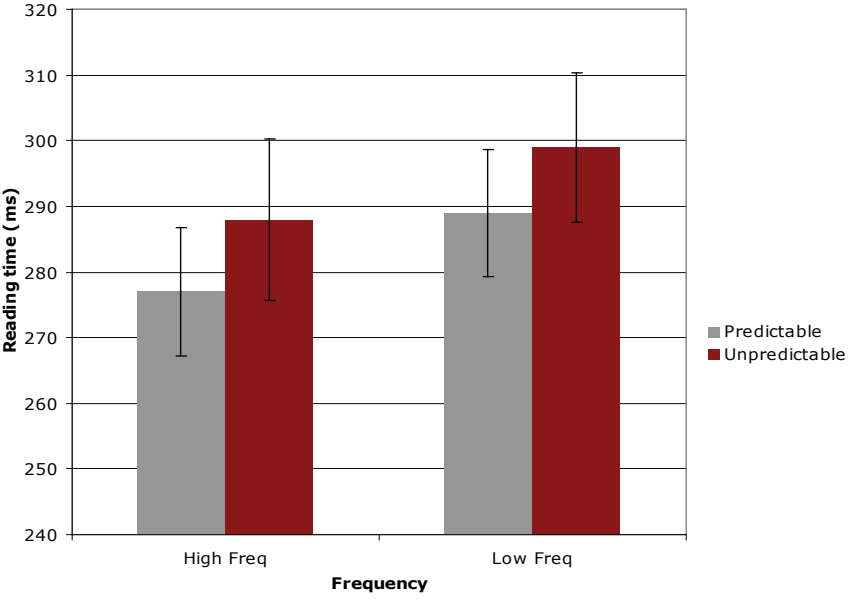


Figure 4. Reading times for target words in Set III as a function of word frequency and predictability. Both variables showed main effects, but no interaction.

= .373, $p < .05$). These results are partially consistent with results from experiments using words in isolation, such as with a lexical decision task or a naming task (Nam et al., 1997) both of which showed the effects of other sublexical units (i.e., phonemes, letter) as well. Thus, the sublexical unit effect might be different depending on the task, for example, when native speakers of Korean read a sentence, they may be sensitive to the internal structure, but not phonemes and letters. The task in this experiment was self-paced reading, word by word, and silent reading might not require explicit phonological processing as single word metalinguistic tasks often do.

4.3 Set III: Word frequency and predictability

In Set III, 1.7% of the data in excess of 3 *SD*'s of the mean were excluded from analyses. A 2 (high vs. low frequency) × 2 (predictable vs. unpredictable) ANOVA indicated reliable effects of word frequency [F(1,

40) = 5.54, $p < .01$] and a marginal effect of predictability [$F(1, 40) = 3.80$, $p = .058$], but no interaction ($F < .5$). A power analysis indicated that the latter was not likely to be a Type II error (power = .055).

The overall findings from Set III were thus consistent with previous eye-tracking experiments that have examined the same variables in English (Altarriba, Kroll, Scholl, & Rayner, 1996; Hulk et al., 2017; Rayner, Binder, Ashby, & Pollatsek, 2001; Rayner et al., 2004) or French (Miellet et al., 2007). As in these experiments that used English, Korean also showed the effects of word frequency and predictability, but no interaction. Therefore, we can conclude that word frequency and predictability affect lexical access additively or that the effect different stages of lexical processing (e.g., pre-lexical vs. post-lexical processing). The fact that there was a relatively weak effect of word predictability in Korean can probably be explained by the relatively lower cloze probability in this experiment.

5. General Discussion

In the present study, a self-paced reading task was administered to examine how three types of factors, visual, lexical, and contextual factors, affect word recognition during reading Korean sentence. The results showed that reading times of Korean words within a sentence were affected by the number of visual features and the number of syllables, but not by number of phonemes or letters. Word frequency, representing a lexical variable, also had a significant effect on reading time for target words as many other previous studies have shown; therefore, we can conclude that readers rapidly access the meaning of frequent words as compared to infrequent words. In addition, the interaction between frequency and the number of syllables was significant. Another major finding of the present study was that word predictability reliably affected the reading times of target words. However, an interaction between word frequency and word predictability was not found.

Results of the present experiment can provide an opportunity to compare word recognition mechanisms in Korean with findings from studies using other languages. By comparing empirical findings from different languages, we can delineate the universal properties of word recognition and language

dependent processing in word recognition as well.

Perfetti (2003) argued that reading has universal properties that reading requires the reader to make links to language at the phonological and morphemic levels. At the same time, the nature of the writing systems and orthographies that instantiate them do make a difference for important details of the reading process. In terms of universality in reading, we also speculate that word recognition is the foundation of skilled reading (Gough, 1984), focused on its universal properties and language-dependent constraints.

To address comparisons of Korean and other languages, we focus on only a few highly distinct languages that represent different writing systems. Languages may be distinguished from one another via their writing systems and orthographies. As Perfetti (2003) illustrated, writing systems can be understood at two levels; mapping principle as a higher level and orthographic constraint as a lower level. Writing systems can be generally classified in three ways: Alphabetic, Syllabic, and Logographic. In addition, one of the approaches to an individual writing system could be the analysis of its orthography, and this approach might be useful to observe language dependent processing.

5.1 Processing Unit and Writing System

One of the goals of the present experiment was to examine what reading units have a significant role in word recognition in Korean. Among the various reading units, what types of units seriously affect word recognition process, and what is the preferred processing unit depends upon the type of writing system? According to Taylor and Taylor (1983), various definitions of a reading unit are possible on the broad level. For example, a reading unit could be physically separable from others of its kind; it may have a sound; it may have a meaning. Several types of units can be summarized to two different types of processing units in a word. The first is the meaningless or sublexical units, such as letters, letter clusters, and syllables. The syllable is the unit of writing in a syllabary, such as the Japanese Kana, and syllables represent morphemes in the Chinese language and writing. The second type of units are meaningful or lexical units which are morphemes. Morphemes are the smallest meaningful units, and they may consist of a few alphabetic

letters or one Chinese character. Stems, as free morphemes, and affixes, as bound morphemes, are also recognized as morpheme units.

Korean, having an alphabetic syllabary, shares properties with both English (alphabetic) and Chinese (morpho-syllabic). This property is shown in the processing unit of Korean. First, as compared to the Chinese writing system, both Korean and Chinese have a similar written form - square-shaped characters. In particular, Korean words of the CVCC type are more similar to Chinese logographs (Taylor, 1980). Although the internal structure variables are not directly related to word reading units, their influence has been reported. With respect to the internal structure, visual properties might play a role in both Chinese and Korean. In Chinese, there is a complexity of character which is measured by the number of strokes (Taylor, 1980). A stroke is a dot, an L-shape, or a horizontal, vertical, or diagonal line, and there are about 20 stroke types (Wang, 1981). A stroke in Chinese character is very similar to a visual feature in Korean (Nam et al., 1997). Neither stroke nor visual feature relates any lexical or semantic information in Chinese and Korean. But, Yeh and Liu (1972) reported an adverse effect of complexity on recognition: latency for recognition was longer for complex characters (15 or more strokes) than for simple ones (10 or fewer strokes). A similar effect was found in a previous study of Korean using a lexical decision task (Nam et al., 1997) and the current experiment using self-paced reading task supports this conclusion. These results support the conclusion that both native speakers of Chinese and Korean are sensitive to this sort of visual factor, even if it is not directly related to word reading units.

Next, considering a purely alphabetic language, such as English, results from experiments showed the effect of word length, which can be measured in the number of letters (Lee, 1999; Balota & Chumbley, 1984) or in the number of syllables (Lovatt, Avons, & Masterson, 2000). These results indicate that short words (e.g., one syllable) gain a benefit not only from working memory span (Baddeley, Thomson, & Buchanan, 1975), but also from their decreased load on the phonological memory system (Lovatt et al., 2000).

In Korean, previous studies (Lee & Kim, 1989; Nam et al., 1997) speculated that the number of syllables would potentially affect the latencies for naming or lexical decision tasks. In addition, Nam et al.

(1997) found effects based on other sublexical units, such as the number of letters or phonemes. In contrast to the experiments using naming or lexical decision tasks, the present study employed self-paced reading and only found an effect of the number of syllables. Therefore, considering both the characteristics of the Korean writing system (i.e., alphabetic syllabary) and task difference (self-paced reading), these findings suggest that the preferred processing unit of printed Korean words is the syllable, which takes the internal structure of Hangul variation into account. This conclusion is also consistent with some previous results which found that the frequency of syllable and the number of syllables that were kulja (Yi, 1993; Lee & Kim, 1989).

5.2 Word recognition within sentences

The task situation in this experiment, words within a sentence, was reflected in some of the results. The first relevant result is reading time across words with different numbers of syllables, which was linearly increased as the number of syllables increased (i.e., 1-4 syllables). This is inconsistent with the results from the previous studies of word recognition in Korean (Park, 1993; Bae, Park, Lee, & Yi, 2016). For instance, Park (1993) showed that latency for one-syllable words in lexical decision is much longer than two-syllable words (i.e., monosyllabic word inferiority effect). Park interpreted this result to mean that monosyllable words have greater semantic uncertainty compared to words with different numbers of syllables. However, the design of the current study, where a word was embedded in a sentence, could eliminate the semantic uncertainty of monosyllable words, so the reading time for monosyllable words have been found to be faster than two-syllable words.

Another result related to having the word in a sentence is that the current experiment did not find the effects of the number of letters or phonemes. A previous study (Nam et al., 1997) found both main effects in naming and lexical decision. It is possible that the effects of both sublexical units can be reduced in words in text. Silent reading also implies the inner articulation of words. It is also possible, however, to diminish the resource for referring to the phonemic information to process word meaning. In other words, readers are expected to not necessarily access all of phonological information

from the printed words during silent reading task. In addition, we also can consider a relationship between task characteristics and orthographic depth. This effect could be more salient in shallow orthographic scripts (i.e., Korean), and, as a result, there could be no effect of the number of phonemes as seen in this experiment, differently from Nam et al. (1997)'s findings.

Next, interaction between word length and frequency can be discussed. Regression analyses from Set I showed an interaction between the number of syllables and frequency. In regards to this interaction, the frequency effects were larger in one-and four-syllable words than in two-and three-syllable words. This pattern of interaction reflects the existence of syllabic types (1–4 syllable) in the corpus that was referred to for selecting stimuli. The ratio of one-syllable word: two-syllable word: three-syllable word: four-syllable word in the corpus is to 1: 17: 15: 4. Therefore, it is plausible that the processing latency for one-and four-syllable words is more sensitive to word frequency than for two-and three-syllable words. This is possible because native Korean readers have more experience with two-and three-syllable words than with one-and four-syllable words in their reading situations, and these different experience may influence Korean readers' different sensitivity to word frequency depending on the number of syllables. Note that this tendency of interaction was different in previous studies (Nam et al, 1997; Park, 1993).

5.3 Context Effect: Word Predictability

The strength of sentential constraints was normalized as the predictability of target words, and the cloze task conducted relatively low cloze probability probably due to the head-final structure of Korean sentences. The result showed, however, a significant but a slightly smaller effect of predictability compared to English studies. The effect of word predictability could provide additional information about a dimension of word recognition, namely word in text.

Next, the question about interaction between the predictability effect and frequency effect arises. First of all, a word frequency effect was shown robustly, implying that we probably have easier access to the mental lexicon in frequent words than in infrequent words. Interestingly, however, there

was no interaction between these two variables in the present experiment. This is consistent with several eye tracking studies (Miellet et al., 2007; Hulk et al., 2017; Rayner et al., 2004). It might be explained by logic of interaction patterns (Sternberg, 1969). According to Sternberg, if there is an interaction between two factors, both factors affect cognitive processes at the same stage. Otherwise, if there is no interaction, both factors share a processing stage at least.

Finally, it is deserved to discuss how the three main variables of interest (word length, frequency, and predictability) affect the reading of Korean. Based on all of the findings from the current study, a comprehensive understanding of the reading of Korean could be described in the following. In the early stage of reading Korean, word length (i.e., syllable) and visual complexity (or perceptual difficulty) influence word recognition. At the same time, word predictability, which is heavily constrained by the prior syntactic and/or semantic context, also impacts this process. This is plausible because a previous study using Korean (Lee, 2004) showed a reliable interaction between predictability and visual gradation. Therefore, we can conclude that predictability is also involved, at least, in the early stage of processing with Korean. After this prelexical processing, a reader acquires the word meaning by accessing the lexical information of the given words. At this stage, the word-frequency effect plays a critical role for achieving this process, and the degree of the frequency effect varies with the word length. Again, the word-length effect lasts from the prelexical processing stage to the lexical processing stage.

6. Conclusion

The evidence from the present study has some implications. As an alphabetic syllabary writing system, Korean word recognition is reliably affected by the number of syllables and visual features. It indicates that Korean writing system shares its properties with neighbor writing system (alphabetic, syllabic, and logographic). Therefore, both visual features as a logographic property and syllables as an alphabetic property have influence on Korean word recognition. Although the current study could not explain why there were no effects of letters and phonemes on reading times, it was

possible that the task situation (i.e., word in sentence) may contribute to the reduced effect of the sublexical units. Word frequency is a strong factor for Korean word recognition regardless of task. Word predictability also has a significant role in this process, even if its cloze probability was lower (by weak syntactic constraints) than English study. This may reflect a broad difference between languages: Korean (or Chinese) is more topic prominent language rather than English which is more syntax prominent language (Li, 1976).

Finally, better understanding about the relationship between language-specific details and task characteristics, by the systematic manipulations, will provide more accurate assessments of word recognition processes.

References

Allen, P. A., McNeal, M., & Kvak, D. (1992). Perhaps the lexicon is coded as a function of word frequency. *Journal of Memory and Language, 31*, 826-844.

Altarriba, J., Kroll, J. F., Scholl, A., & Rayner, K. (1996). The influence of lexical and conceptual constraints on reading mixed language sentences: Evidence from eye fixation and naming times. *Memory and Cognition, 24*, 477-492.

Bae, S., Park, K., Lee, H-W., & Yi, K. (2016). The mono-syllabic word inferiority effect within Korean word recognition. *The Journal of Linguistic Science, 77*, 109-125.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory, *Journal of Verbal Learning and Verbal Behavior, 14,* 575-589.

Balota, D. A. & Chumbley, J. I. (1984) 'Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage', *Journal of Experimental Psychology: Human Perception and Performance* 10, 340–357

Balota D. A., & Chumbley, J. I. (1985). The locus of word-frequency effects in the pronunciation task: Lexical access and/or production? *Journal of Memory and Language. 24*, 89-106.

Balota D. A., & Yap, M. J. (2006). Attentional control and flexible lexical processing: Explorations of the magic moment of word recognition. In S. Andrews (Ed.), *From Inkmarks to Ideas: Current Issues in Lexical Processing*, 229-258.

Frost, R., Katz, L., & Bentin, S. (1987). Strategies for visual word recognition and

orthographical depth: A multilingual comparison. *Journal of Experimental Psychology: Human Perception and Performance, 13*, 104-115.

Gough, P. B. (1984). Word recognition. In P.D. Pearson, R. Barr, M.L. Kamil, & P. Mosenthal (Eds.), *Handbook of reading research* (pp. 225-254). White Plains, NY: Longman.

Huck, A., Thompson, R. L., Cruice, M., & Marshall, J. (in press). Effects of word frequency and contextual predictability on sentence reading in aphasia: an eye movement analysis. *Aphasiology.*

Inoue, A. (1991). A comparative study of parsing in English and Japanese. Doctoral dissertation, University of Connecticut , Storrs.

Juphard, A., Carbonnel, S., & Valdois, S. (2004). Length effect in reading and lexical decision: Evidence from skilled readers and a developmental dyslexic participant. *Brain and Cognition, 55*, 332-340.

Kim, S. Y., & Bolger, D. J. (2016). The role of subsyllabic units in the visual word recognition of Korean monosyllabic words: A masked priming study. *Journal of Cognitive Science, 17*, 343-359.

Lavidor, M., & Whitney, C. (2005). Word length effects in Hebrew. *Cognitive Brain Research, 24,* 127-132.

Lee, C. H. (1999). A locus of the word-length effect on word recognition. *Journal of Reading Psychology, 20*, 129-150.

Lee, H. (2004). The influences of context on visually-degraded word recognition in Hangul reading. *The Korean Journal of Experimental and Cognitive Psychology, 16*, 467-482.

Lee, J.-S., & Kim, K.-L. (1989). The processing unit of Korean Words. *Korean Journal of Cognitive Science, 1*, 221-239.

Li, C. N. (1976). *Subject and Topic.* New York: Academic Press.

Lima, S. D., & Pollatsek, A. (1983). Lexical access via orthographic code?: The Basic Orthographic Syllabic Structure (BOSS) reconsidered. *Journal of Verbal Learning and Verbal Behavior, 22*, 310-332.

Lovatt, P., Avons, S. E., & Masterson, J. (2000). The word length effect and disyllabic words. *Quarterly Journal of Experimental Psychology, 53*, 1-22.

Mazuka, R., & Lust, B. (1990). On parameter-setting and parsing: Predictions for acquisition. In L. Frazier & J. de Villers (Eds.), *Language processing and acquisition* (pp. 163-206). Dordrecht: Kluwer.

Monsell, S., Doyle, M. C., & Haggard, P. N. (1989). Effects of frequency on visual word recognition tasks: Where are they? *Journal of Experimental Psychology: General. 118*, 43-71.

Miellet S., Sparrow, L., & Sereno, S. C. (2007). Word frequency and predictability effects in reading French: An evaluation of the E-Z Reader model. *Psychonomic Bulletin and Review, 14*, 762-769.

Nam, K., Seo, K., Choi, K., Lee, K., Kim, T., & Lee, M. (1997). The word length effect on Hangul word recognition. *The Korean Journal of Experimental and Cognitive Psychology, 9*, 1-18.

National Academy of the Korean Language (2002). *The Frequency Database of Korean Words. http://www.korean.go.kr*

New, B., Ferrand, L., Pallier, C., & Brysbaert, M. (2006). Re-examining the word length in visual word recognition: new evidence from the English Lexicon Project. *Psychonomic Bulletin and Review, 13*, 45-52.

Park, K. (1993). Mental code involved in Hangul word recognition. *The Korean Journal of Experimental and Cognitive Psychology, 5*, 40-55.

Park, K. (1996). The role of phonology in Hangul word recognition. *The Korean Journal of Experimental and Cognitive Psychology, 8*, 25-45.

Perfetti, C. A. (2003). The universal grammar of reading. *Scientific Studies of Reading, 7*, 3–24.

Prinzmetal, W., Treiman, R., & Rho, S. H. (1986). How to see a reading unit. *Journal of Memory and Language, 25*, 461-475.

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

Rayner, K., Ashby, J., Pollatsek, A., & Reichle, E. D. (2004). The effects of frequency and predictability on eye fixations in reading: Implications for the E-Z Reader Model. *Journal of Experimental Psychology: Human Perception and Performance, 30*, 720-732.

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

Spoehr, K. T., & Smith, E. E. (1973). The role of syllables in perceptual processing. *Cognitive Psychology*, 5, 71-89.

Stanovich, K. E., & West, R. F. (1983). On priming by a sentence context. *Journal of Experimental Psychology: General, 112,* 1-36.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. In W. G. Koster (Ed.), *Attention and Performance II*. Amsterdam: North Holland.

Taylor, I. (1980). The Korean writing system: An alphabet? A syllabary? A logography? In P. A. Kolers, M. E. Wrolstad, & H. Bouma (Eds.), *Processing of visible language (Vol. 2)*. (pp. 67-82) New York: Plenum.

Taylor, I., & Taylor, M. M. (1983). *The psychology of reading*. NEW YORK: Academic Press

Tsai, C.-H. (2001). *Word identification and eye movements in reading Chinese: A modeling approach*. Doctoral dissertation, University of Illinois at Urbana-

Champaign.
Wang, S-Y. W. (1981). Language structure and optimal orthography. O. J. L. Tzeng & H. Singer (Eds.), *Perception of print: reading research in experimental psychology*. (pp. 223-234). Hillsdale, NJ: Lawrence Erlbaum Associates.
Yeh, J.-S., & Liu, I.-M. (1972). Factors affecting recognition thresholds of Chinese characters. *Acta Psychologica Taiwanica, 14*, 113-117.
Yi, K-O., (1993). Effect of orthography on mental lexicon structure and search during the Korean word identification. *The Korean Journal of Experimental and Cognitive Psychology, 5*, 26-39.

Appendix

The sentence stimuli used in the Experiment are shown below. English translations for each word and /or phrase are presented below each sentence.

Stimuli for Set I: 1–4 Syllable Words with Case Marker Printed Italics are Target Words.

유럽 방문을 / 무사히 마치고 / 돌아온 / *대통령이* / 환영을 받았다.
Europe visiting / safely finishing / come back / *president* / welcomed

진용은 / 자신의 미래를 / 고민하다가 / 마침내 / *선생님이* / 되기로 / 결심했다.
Jin-Yong / his future / thinking about / finally / *teacher* / becoming to / decided

기업이 / 성공하기 위해서는 / 무엇보다 / *소비자의* / 욕구를 / 파악해야 한다.
A company / to success / most of all / *customer* / needs / have to know

신영이는 / 하루종일 / 피곤하셨을 / 엄마를 위해서/ *발을* / 주물러 드렸다.
Shin-Young / all day long / should be tired of / to mother/ *foot* / massaged

승용이는 / 엄마를 닮아서인지 / 음식을 먹을 때 / *땀을* / 많이 흘린다.
Seung-Yong / mother same as / when eat food / *sweat* / too much

지영이는 / 스무 살이 넘도록 / *비행기를* / 타보질 못했다.
Ji-Young / twenty years over / *airplane* / take did not/

열심히 운동하는 / 도연이도 / 나이가 드니 / 별수없이 / *배가* / 나오기 시작한다.
Hard working out / Do-Youn / getting old/ unavoidable / *belly* / develop start

오늘 / 아침 밥상에는 / 예상치 못했던 / *미역국이* / 올라와 있었다.

Today / morning breakfast / unexpected / *Mi-Yek soup* / was there

어릴 때와는 달리 / 조카는 / 이제 잠자기 전에는 / *양치질을* / 하는 / 습관이 생겼다.
Childhood different / nephew / now go to bed before / *brushing* / do / habit accustomed

우리나라의 산이 / 아름다운 / 또 다른 이유는 / 수많은 / *문화재를* / 보유하고/ 있기 때문이다.
Our country's mountain/ beautiful/ another reason/ lots of/ *cultural-assets*/ having / because

아침에 / 일어나보니 / 지난 밤 / 내내 / *함박눈이* / 내린 것을 / 알 수 있었다.
Morning / get up / last night / during / *snowflake* / coming down / was able to know

영식이는 / 어머니 생신을 위해서 / 좋아하시는 / *제비꽃을* / 선물로 / 골랐다.
Young-Shik / mother birthday to / like / *violet* / gift / chose

그 엄청난 규모의 도시를 / 한마디로 표현하자면 / *교통지옥이* / 가장 잘 어울릴 것이다.
The tremendous scale city/ one word expression/ *traffic-jammed-hell* / most plausible

먼 조상이 만들었다는 / 그 도자기가 / 집안의 / *가보로* / 전해져 내려온다.
Ancestry made / the pottery/ family/ *heirloom* / handed down

인수는 / 대학을 졸업한 / 늦은 나이에 / *담배를* / 피우기 시작했다.
In-Soo / college graduated / old age / *cigarette* / smoking started

광식이의 취미는/ 호숫가에서/ 한가롭게/ *산책을* / 즐기는 것이다.
Kwang-Shik's hobby/ in the lake/ leisurely/ *take-walk* / enjoying is

영어가 익숙해진/ 수진이는/ 이제 / *일상용어를* / 쓰는데 / 문제가 / 없다.
English accustomed to/ Soo-Jin/ now/ *everyday words* / to use / problem /

not

언제나 그런 것은 아니지만/ *뜬소문이* / 사실일 수도 있다.
Always not though / *unfounded-rumor* / true possibly

나의 어린시절을 / 떠올리면 / 언제나 / *할아버지의* / 추억이 / 먼저 생각난다.
My childhood / remembering / always / *grandfather* / memory / first come

내가 무슨 잘못을 하든지 간에 / 그것은/ 집안에 / *먹칠을* / 하는 것이다.
Whatever I did wrong/that/ family / *dishonor* / to do

운동을 하기로/ 결심한 이후에/ 지금까지/ *산에* /오르는 것을/ 계속하고 있다.
Working out planned/ decide after/ so far/ *mountain* / climbing/ keep continuing

꾸준히 하다 보면/ 언제가/ 나에게도/ *빛이* / 보이지 않을까 한다.
Working hard/ sometime/ even for me/ *light* / can be seen

그 기업의/ 판매전략은/ 소비자들의 자발적인/ *불매운동을* /부추기고 있다.
That company/ marketing strategy/ customers' voluntary/ *boycott* / instigating

옛날에는/ 마을에 급한 일이 생기면/ *종을*/ 울려서 사람들에게 알리기도 했다.
Long time ago/ village emergency happen/ *bell*/ ringing to inform people

산에 갔다가 / 갑작스럽게 날아든 / *벌에* / 깜짝 놀랐다.
Mountain climbing/ suddenly attack/ *bee* / shocked

긴 여행을 마친 / 그 여객선이/ *닻을* / 내리고/ 항구에 정박했다.
Long travel finishing/ the passenger-ship/ *anchor* / down/ in the harbor

마음이 고운/ 그 아이는/ 친구의/ *꾐에* / 빠지고 말았다.
Warmhearted/ the boy/ friend's/ temptation / into the way

선경이는/ 여행갈 때는/ 언제나/ *빗을* / 반드시 챙긴다.

Sun-Kyung / when traveling / always / *comb* / necessarily check

그는 언제나 자신의 미래를/ 생각할 때/ *백만장자를* / 꿈꾸곤 했다.
He always his future/ think about/ *millionaire*/ dreamed

부모는/ 자식을/ 가르치는데 있어서/ *독립심을* / 키워줄 수 있도록 해야 한다.
Parents/ child/ educate something / *independence* / raise have to

하루종일/ 전시회 구경을 하고 나왔는데/ 어떤 / *작품도* / 기억 나질 않는다.
All day long/ exhibition seeing/ any/ *work*/ cannot remember

사회가 급속도로 서구화되면서/ 우리의/ *대중문화도* / 상당히 변질되었다.
Society rapidly Westernization/ our / *popular-culture* / considerably changed

Stimuli for Set II: 2-syllable Words with Case Marker Printed Italics are Target Words. The Number of Phonemes, Letters, and the Number of Visual Features are Varied.

서울의 여름이 괴로운 이유는 / 무엇보다 높은 / *습도에* / 있다고 할 수 있다.
Seoul summer painful reason/ most of all high / *humidity* / there is

시민단체는/ 기업들의/ 담합행위에 대한/ *반대를* /분명히 표명했다.
The citizen group/ enterprises/ conferring acting/ *objection*/ clearly pronounced

교수의/ 사정에 의해서/ 강의 계획에는/ 약간의 / *차질이* / 생기고 말았다.
Professor/ personal commitment/ class schedule/ slight / *problem* / happened

저녁 식사 후에는/ 남산타워에/ 올라가/ 멋진/ *야경을* 보여줄 것이다.
After dinner/ Nam-San tower/ go to/ fantastic/ *night-view*/ will show

지난밤에/ 돼지 꿈을 꿨으니/ *복권을* / 사야겠다.

Last night/ good dreamed/ *lottery* / should buy

부지런한 동생은/ 아침마다 오는/ *신문을* /맡아서 챙겨둔다.
Industrious brother/ every morning/ newspaper / in charged of

소풍은/ 날씨가 나빠졌기/ 때문에/ 결국/ *취소가* / 되고 말았다.
Picnic/ weather got bad/ because/ finally/ *cancel* / resulted

현준은/ 어느새/ 머리가 희긋한/ *중년이* / 되어 있었다.
Hyun-Joon/ without awareness/ hair grey/ *middle-age* / became

군것질을/ 좋아하는/ 수빈이는/ 영화를 볼 때 주로/ *팝콘을* / 먹는다.
Candy/ like/ Soo-bin/ during watch movie usually / *popcorn* / eat

유원이는/ 배고플 때/ 집안에/ 밥이 없으면/ 주로/ *라면을* / 먹는다.
Yoo-won/ hungry/ home/ rice empty/ usually/ *Ramen*/ eat

용욱이는 밤새워 시험 공부를/ 하기 위해서/ 한잔의/ *커피를* / 마시기로 했다.
Yong-Wook/ overnight exam studying/ to do/ a cup of / *coffee* / drink

강의는/ 교수의/ 학회 참석의 사정으로/ 인해서/ *휴강이* / 되고 말았다.
Class/ Professor/ conference visiting reason/ by/ *no-class* / resulted

그나마/ 다행인 것은/ 이런 고민을 털어놓을/ *친구가* / 있다는 것이다.
Nevertheless/ lucky thing/ this anguish confess/ *friend*/ there is

요새/ 화제가/ 되고 있는/ 그 영화의 소재는/ 알고 보니/ *실화로* / 밝혀져 충격을 주었다.
Recently/ topic/ becoming/ the movie material/ revealed/ *true-story*/ discovered/ shocked

각자의 생활이 바쁘다 보니/ 온 가족이 함께 아침마다/ *식사를* / 하기가 어렵다.
Individual life busy/ all family together every morning/ *breakfast* / doing difficult

밖에/ 나가보니/ 비가 오길래/ 다시 들어가/ *우산을* / 챙겨서 나왔다.
Outside/ go to/ raining/ again come in/ umbrella / taking go out

집에/ 와보니/ 우체통에/ 친구가 보낸 선물/ *상자가* / 들어 있었다.
Home/ back/ mailbox/ friend sent present/ box / there was

불우 이웃 돕기 행사에/ 시민들의 자발적인/ *참여가* / 계속해서 이어졌다.
Campaign to help unfortunate/ the citizens voluntary / *participation*/ continuously went

군대에 있는 친구를 위해서 준비한/ 선물을/ *소포로* /보내기로 했다.
Prepared to friend in army/ present/ *parcel* / sent

용석이는/ 어른이 되면/ 아픈 사람을 돕는/ *의사가* / 되고 싶어했다.
Yong-seok / after grown-up/ sick people help / *doctor* / want to be

새로운 경험을 추구하는 사람들의/ 아프리카 / *탐험이* / 유행처럼 번지고 있다.
People who eager to new experience/ Africa / *exploration*/ become to be fashioned

많은 사람들이/ 인구 과밀의/ 서울을 벗어나 / *교외로* / 주거지를 옮긴다.
Many people/ big population/ escape from Seoul/ *suburb*/ move to

그 정치범은 결국/ 생명을 걸고/ 제3국으로 / *망명을* / 시도하게 되었다.
The political prisoner finally/ at the risk of his life/ to the third nation/ *exile*/ tried

어머니는/ 자식의 성공이 대견한 나머지 / *눈물을* / 흘리고 말았다.
Mother/ her son's success/ so proud of / *tear*/ downed

살면서/ 공부하는 것도 중요하지만/ 다양한/ *경험을* / 쌓는 것도 중요하다.
Everyday life/ studying hard important/ various / *experiences* / important too

술을 마시다가/ 시간이 늦어져서 가까스로/ *막차를* / 타고 돌아왔다.

Because of drinking/ time is late / *the-last-bus* / came back

부동산 의혹이/ 제기된/ 그 장관은/ 결국/ *사임을* / 밝히고/ 물러났다.
Illegal real estate rumor/ proposed/ the minister/ finally / *resignation* / expressed/ left

희정이는/ 스무 번째 생일도 되었으니/ 이제/ 어엿한 / *숙녀가* / 된 기분이 들었다.
Hee-jeong/ twenty year's birthday/ now / *lady* / felt / to be

앉는 자세가/ 삐딱하면/ 흔히/ *척추에* /문제가/ 생기기 쉽다.
Seating posture/ not good/ commonly/ *backbone*/ problem/ occurs easily.

의대에서/ 유급 당하지 않으려면/ 한과목이라도/ *낙제를* / 당해서는 안 된다.
Medical school/ not to remain/ even one subject / *failure* / not allowed

승식이는 사춘기가 되면서/ 여드름이/ 잔뜩 난/ *피부가* /고민이다.
Seung-shik/ becoming adolescence/ pimples/ many /*skin* / concerns

강한 햇빛을/ 너무 오래/ 쬐는 것은/ *얼굴에* /좋지 않다.
Strong sunshine/ too long/ taking / *face* / not good

Stimuli for Set III: Low-Frequency Predictable (LFP) or High-Frequency Unpredictable (HFUP)

최근 들어/ 극장가에서는/ 한국영화가/ 기록적인 [*흥행을/인기를*] 끌고 있다.
Recently/ film market/ Korean movie/ record-breaking [*sensation/ popularity*] catching

구원투수가/ 역전을 허용하자/ 홈팀 관중들의 [*야유가/분노가*] 쏟아져 나왔다.
Relief/ allow reverse match/ home team fans [*ridicule/anger*] poured

차가 더러워졌으니/ 일요일에는/ 반드시 [*세차를/청소를*] 해야겠다.
Car got dirty/ on Sunday/ necessarily [*car-wash / cleaning*] have to

선생님께/ 칭찬 받으려고/ 다음주 수업을 [*예습/공부*] 하기로 했다.
From teacher/ applauded/ next week class [*preview/study*] decided to do

시험기간만 되면/ 평소에는 텅텅 비었던 [*도서관이/고시원이*] 학생들로 꽉 찬다.
During exams/ usually empty [*library/examination-room*] students/ full of

오늘도/ 늦는 걸 보니 남편은/ 또/ 직원들과 [*회식이/모임이*] 있는 것 같다.
Today/ late husband/ again/ with staffs [*mess/meeting*] likely to be

그 회사는/ 간신히 부도의 위기를 [*모면/해결*] 할 수 있었다.
The company/ hardly/ dishonor crisis [*evade/solve*] was able to

엄마는/ 아빠의 숙취/ 해소를/ 위해서/ 한잔의 [*꿀물을/우유를*] 드리기도 한다.
Mom/ dad's hangover/ resolve/ in order to/ a glass of [*honeyed-water/milk*] serve

지난해/ 성적이 좋았던/ 야구 선수에게/ 높은 금액의 [*연봉이/임금이*] 약속될 것이다.
Last year/ record good/ to baseball player/ great amount of [*annual-salary/wages*] will be promised

선거가/ 불리해 지자/ 상대편 후보에 대해/ 끊임없는 [*비방을/공격을*] 멈추지 않았다.
Election/ disadvantageous/ to the other candidate/ ceaseless [*abuse/attack*] do not stop

그 펀드매니저는/ 자신의 퇴직금을/ 모두 [*기부할/투자할*] 곳을 결정했다.
The fund manager/ ones' retirement pay/ all [*donation/investment*] decided

이번 대회의/ 좋은 성적은/ 알고 보니/ 꾸준한 [*연습을/합숙을*] 했던 결과였다.
This series/ good record/ discovered/ industrious [*training/ joint-billet*] resulted by

방학이 되면/ 그 동안 가보고 싶었던/ 바닷가로/ [*여행을/일출을*] 즐기러/ 갈 것이다.

Vacation comes/ wanted to go/ the seaside / [*trip/sunrise*] to enjoy / will go

최하위였던 팀이/ 우승 후보를/ 꺽는/ 놀라운 [*이변이/결과가*] 다시 한번 나타났다.
The lowest ranked team/ champion favorite/ beating/ surprising [*extraordinary-event/result*] once again/ shown

혼자서 오래 자취생활을 하다 보니/ 남자이지만 [*요리에/취사에*] 자신이 있다.
Living alone for a long time/ even man [*cooking/kitchen-work*] be good at

자기 마음에 들지 않는다고 해서 / 이러쿵 저러쿵 [*불평은/판단은*] 좋지 않다.
Even if you don't like something/ this and that-like [*complain/judge*] not good

Stimuli for Set III: High-Frequency Predictable (HFP) or Low-Frequency Unpredictable (LFUP)

유행에 민감한 엄마는/ 요즘엔/ 어떤 배우가 /가장 [*인기를/흥행을*] 끌고 있는지 알고 계신다.
Fashioned sensitive mom/ recent/ which actor/ the most [*popularity/sensation*] be knowing

사회 저명인사의/ 친일 행적에 대해서/ 국민들의/ [*분노가/야유가*] 쏟아져 나왔다.
Society celebrity/ to pro-Japanese/ people [*anger/ridicule*] poured

일요일에는/ 그 동안/ 미뤘던/ 빨래와/ [*청소를/세차를*] 해야겠다.
On Sunday/ for a while/ postponed/ washing and [*cleaning/car-wash*] have to

고 3이 된/ 은희는/ 대학에 가기 위해서/ 열심히/ [*공부/예습*] 하기로 했다.
High school 3 year/ Eun-Hui/ to enter college/ hard [*study/preview*] decided

사법시험을/ 준비하는 사람들은/ 학원가/ 근처의/ [*고시원이/도서관이*]

공부하기 좋다.
Law examination/ preparing people/ institute-street/ nearby [*examination-room/library*] good for studying

한 달에 한번씩은/ 어릴 적 친구들과/ [*모임이/회식이*] 있는 것 같다.
Once a month/ with old friends [*meeting/mess*] likely to be

가끔/ 전문가에게/ 조언을/ 구하는 것이/ 문제를 [해결/모면] 할 수 있는 길이 되기도 한다.
Sometime/ expert/ advice/ to ask/ problem [*solve/evade*] can be a way

나는/ 아침을/ 안 먹는/ 대신에/ 한잔의 [*우유를/꿀물을*] 마시기도 한다.
I/ breakfast/ not eat/ instead of/ a glass of [*milk* / *honeyed-water*] drink

회사측은/ 노조의/ 입장을/ 받아들여서/ 내년부터 [*임금을/연봉을*] 올려주기로 했다.
Company/ union/ opinion/ accept/ next year/ [*wages/ annual-salary*] will increase

미국과 영국 군대는/ 이라크에 대한/ 끊임없는 [*공격을/비방을*] 멈추지 않았다.
U.S and British army/ to Iraq / ceaseless [*attack/abuse*] never stop

신기술 개발 경쟁에서/ 앞서기 위해서/ 기업들은 막대한 [*투자를/기부를*] 하고 있다.
New technology development competition/ to advance/ company/ great [*investment/donation*] doing

국가 대표 선수들이/ 대회를/ 앞두고/ 3개월 간의 [합숙을/연습을] 했던 것이 효과를 보았다.
National player/ game/ in prior to/ 3 months [*training/ joint-billet*] effect was shown

새해의/ 첫날에는/ 동해에 가서 [*일출을/여행을*] 즐길 것이다.
New year/ first day/ East sea going to go [*sunrise/travel*] will enjoy

그 연구팀의/ 반복된 실험에서/ 예상하지 못한/ 놀라운 [*결과가/이변이*] 다시 한번/ 나타났다.
The research team/ repeated experiments/ unexpected/ surprising [*result/ extraordinary-event*] again/ come up

모든 국립공원은/ 산불을 방지 위해서/ 산에서의 [*취사를/요리를*] 금지하고 있다.
All national parks/ to prevent fire/ in the mountain [*kitchen-work/cooking*] not permitted

잘 모르는 사람에/ 대해서 섣부른 [*판단은*/ 불평은] 좋지 않다.
Unknown person/ about/ early [*judge/complain*] not good