

The exploration of the effects of word frequency and word length on Korean word recognition

Changhwan Lee¹, Yoonhyoung Lee^{2*}, Tae Hoon Kim³

¹Department of Psychology, Sogang University

²Department of Psychology, Yeungnam University

³Department of Psychology, Kyungnam University

한국어 단어재인에 있어서 빈도와 길이 효과 탐색

이창환¹, 이윤형^{2*}, 김태훈³

¹서강대학교 심리학과, ²영남대학교 심리학과, ³경남대학교 심리학과

Abstract Because a word is the basic unit of language processing, studies of the word recognition processing and the variables that contribute to word recognition processing are very important. Word frequency and word length are recognized as important factors on word recognition. This study examined the effects of those two variables on the Korean word recognition processing. In Experiment 1, two types of Hangeul words, pure Hangeul words and Hangeul words with Hanja counterparts, were used to explore the frequency effects. A frequency effect was not observed for Hangeul words with Hanja counterparts. In Experiment 2, the word length was manipulated to determine if the word length effect appears in Hangeul words. Contrary to the expectation, one syllable words were processed more slowly than two syllable words. The possible explanations for these results and future research directions are discussed.

요약 단어는 언어의 기초적인 의미 단위가기 때문에 단어재인에 대한 연구는 언어 연구에서 중요하며 단어처리에 기여하는 변인이 무엇인지에 관한 연구가 이루어져 왔다. 본 연구에서는 한국어 단어재인 과정의 주요 변인 중 단어 빈도와 단어 길이의 영향을 탐색하였다. 먼저 단어 빈도와 관련하여, 한국어의 특징 중 하나인 한자어로 이루어진 단어에서도 기존의 연구와 동일한 양상의 빈도 효과가 나타나는지를 탐색하였다. 이를 위해 순 한글 단어와 한자어로 이루어진 단어를 비교하였으며, 그 결과 한자어로 이루어진 단어에서는 빈도 효과가 나타나지 않았다. 한편 단어 길이 효과의 경우, 단음절로 구성된 단어의 양상을 확인해 보고자, 음절의 개수를 변화시켜 단어 길이 효과를 측정하였다. 그 결과 단음절 단어는 이음절 단어에 비해 느리게 처리되었다. 특정 유형의 단어에 대한 빈도 효과의 부재 및 단음절 단어의 느린 처리는 한국어의 특징을 반영한 결과라 할 수 있으며 추후 연구를 통해 이에 대한 좀더 자세한 탐색이 필요할 것이다.

Keywords : Effect of morphemes, Korean word, Word recognition, Word frequency, Word length

1. Introduction

Frequency and length effects are the main lexical variables in word recognition[1-12]. For example,

deciding the lexicality of a word is easier for high frequency words than for low frequency words in the lexical decision task. Therefore, more frequent words require less time to process than do less frequent

This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2013S1A3A2054928)

This work was partly supported by Sogang University research fund awarded to the 1st author(201210070.01).

*Corresponding Author : Yoonhyoung Lee(Yeungnam University)

Tel: +82-53-810-2231 email: yhlee01@yu.ac.kr

Received September 24, 2015

Revised December 22, 2015

Accepted January 5, 2016

Published January 31, 2016

words. Similarly, the length effect is the main phenomenon in word recognition, such that longer words require more processing time than do shorter words. While such results were replicated in many other languages, it is not sure whether such phenomena can be applied to Korean orthography.

As related with experiment 1, word frequency has a powerful effect on word recognition in many languages [1,2,5,9,10,11]. However, some suggested that if a word is processed phonologically, it is possible to find no frequency effect [11,13]. Therefore, phonological processes would be dominant if a word is processed with almost the same speed across different frequency ranges. This is because the dual route hypothesis assumes that the phonological route, operated by applying the grapheme-phoneme rule, is less sensitive to lexical variables such as frequency and meaning.

Several previous studies about Korean word frequency effects showed that the word is processed more slowly as the frequency goes down, which is consistent with findings of English words [14,15]. However, several other studies also showed no frequency effect at all [16,17]. These inconsistencies require to discover possible confounding variables.

One possible reason for the inconsistent effect is the fact that Korean is byscriptal. Koreans used Hanja, Chinese characters, in their written language before they invented Hangul, presently dominant language, around 1400, A.D. Hanja is similar to Chinese; it is a logography and has orthographic form similar to current Chinese characters. However, unlike other deep orthographies, Hanja has some unique characteristics. First, the relationship between character and pronunciation is almost identical in Hanja. In other words, a character of Hanja has one phonological form in any context like a shallow orthography. Second, Hanja provided source words for Hangul words when Hangul was invented. In other words, in order to make many Hangul words, Hangul borrowed the pronunciation and meaning from Hanja words. Thus, it was suitable to use Hanja and Hangul together in many

old texts. With respect to the possible confounding variable for the inconsistent frequency effects, the Hangul consequently has two types of words: One type of words is composed of a series of morphemes that has come from the meaning and pronunciation of their Hanja counterpart (i.e., two-letter words have two morphemes; three-letter words have three morphemes) (e.g., “국” in “국민” means country). The other type of words is pure Hangul word that does not have any morphological component in its letter strings (e.g., 사랑).

According to these characteristics of Hangul words, the presence of inconsistent frequency effects in the previous studies could be attributed to the different degrees of morphological components within a word. If the frequency effects are different across the two types of Hangul, we need to be cautious in interpreting the frequency effect in Korean words. Additionally, it indicates that the use of one orthography can interact with another orthography from the same language. Specifically, using Hanja would lower the activation level of corresponding Hangul because the phonology and meaning of Hanja and Hangul are the same. This experiment was designed to examine the frequency effect in two subtypes of Hangul using a basic lexical task, the naming task.

Word length also has a powerful influence on word recognition [2-4,6,9,12]. It is well known that more letters, syllables, or phonemes in a word cause processing time to increase.

A pioneering study on the effect of word length as processing time was done by Whaley [12]. He suggests that the three primary indexes of word length are the number of phonemes, the number of letters, and the number of syllables. These variables are highly correlated with one another in lexical decision performance and form a single factor, which explains performance variance to a similar degree as does the factor related to word frequency.

Studies performed after Whaley [12] consistently replicated the word length effect on processing time

[2,7,19,20]. Studies also proposed that the word length effect is greater for poor readers than for skilled readers, which suggests a holistic process in skilled reading [21,22]. There are only few studies on word length effects in other orthographies. However, any unique characteristic of an orthography might produce different patterns of word length effects. Differences between orthographies can shed light on the question of orthographical optimality because an orthography with reduced word length effect is desirable.

2. Experiment 1

Experiment 1 consisted of two sub-studies. In the first, Hangul words that do not have Hanja counterparts were used, and in the second, Hangul words that have Hanja counterparts were used. It is assumed that Hangul words with Hanja counterpart would be processed without relying on phonological route, whereas Hangul words without Hanja counterparts would be processed more dominantly by the phonological route.

2.1 Experiment 1a

Same participants participated in the two sub-experiments for the frequency effect because it was practically difficult to ask Korean native speakers to participate in the United States. The experiment for naming pure Hangul words was performed first. The reason for executing the studies in this order was that the Hangul with Hanja counterparts might influence the processing of the pure Hangul in a mixed presentation.

2.1.1 Method

Participant Forty graduate students enrolled at the University of Texas at Austin participated in this experiments. All were Korean natives. They all had at least an undergraduate degree.

Materials and procedure 48 stimuli were used in each sub-experiment and 10 additional practice stimuli

were selected. Thus, 16 words for each three word frequency condition were selected. The stimuli were selected from the Korean word frequency count [18]: low frequency words were 2 per million, 5 to 7 per million for middle frequency words, and 20 to 50 per million for high frequency words. Each word frequency condition has sixteen stimuli. Also, they were selected to be yoked in frequency across experiments 1a and 1b. In addition, initial phoneme and the number of orthographic strokes were matched with the stimuli in the experiment 1b. Controlling initial phoneme is important in the naming task because some phonemes are not detected on time (e.g., the unvoiced phoneme such as /s/, /p/). Participants were instructed to name a word on a computer screen as quickly and precisely as possible. "Context," microcomputer program written at University of Texas at Austin, was used for stimulus presentation and performance measurement. The latency of initiation of each response was recorded by a voice key. An experimenter sat beside a participant to record pronunciation error.

Statistical analysis SPSS version 21.0 was used to conduct statistical analysis. The independent variable was the word frequency and the dependent variable was the reaction time for lexical decision.

2.1.2 Results

Pronunciation and delay errors were discarded. However, the overall error rates for each condition were less than one percent. These were assumed to be meaningless for the comparison purposes. Median latencies in each of the three conditions were computed for each participant. Means of these medians are presented in Table 1.

Table 1. Mean Naming Latencies, Standard Deviations, and Error rates

Frequency	RT	SD	Errors
High	504	59.7	0.5
Middle	512	65.5	0.4
Low	523	72.5	0.7

Analysis of variance on the reaction time revealed significant effects of word frequency ($F(2, 78) = 11.8, p < .01$). However, the statistical analysis by item did not reveal the significant effects of frequency ($F(2, 30) = 2.43, p = .11$). The pattern of reaction time showed monotonic increase by the frequency such that the reaction time increased monotonically as frequency increased.

2.2 Experiment 1b

The experiment for naming Hangul words with Hanja counterpart was performed. There was a short delay between Experiment 1a and 1b to prevent carryover effects. The experimenter prepared for the second experiment, and it was started as soon as the participant showed readiness.

2.2.1 Method

Participant Same forty participants as in Experiment 1a participated in this experiment.

Materials and procedure Procedure and stimulus selection were the same as in Experiment 1a. Only difference was the types of Hangul words. In this experiment, Hangul words with Hanja counterparts were used.

Statistical analysis The same methods to Experiment 1a were utilized.

2.1.2 Results

Pronunciation and delay errors were discarded. However, the overall error rates for each condition were again less than one percent as in experiment 1a. Median latency in each of the three conditions was computed for each participant. Means of these medians and error rates are presented in Table 2.

Table 2. Mean Naming Latencies, Standard Deviations, and Error rates

Frequency	RT	SD	Errors
High	544	71.5	0.4
Middle	556	72.8	0.4
Low	540	69.3	0.6

Analysis of variance revealed a significant difference among the conditions by subject analysis ($F(2, 78) = 34.5, p < .01$) and by item analysis ($F(2, 30) = 4.0, p < .05$).

In this experiment, Hangul words with Hanja counterparts did not show the monotonic increase as a function of frequency. Instead, latencies showed an inverted V pattern. One possible explanation is that the repetition effect (using both types) by Hanja counterparts is not general for all frequency ranges, but is specific for certain frequency ranges. For example, low frequency words might have more repetition effects than mid-frequency words.

Simpson & Kang [17] obtained no frequency effects, and infer phonological processing for shallow Hangul. The current experiment demonstrated a frequency effects for pure Hangul. This inconsistency could be due to the use of different types of Hangul in Simpson & Kang's study. Specifically, Simpson and Kang used Hangul words with Hanja counterparts. Repetition effects influences the processing of Hangul words that are used in Hanja as well.

This experiment also indicates that the use of another orthography, Hanja, can strengthen the activation level of main orthography, Hangul. Specifically, if a Hanja word which has a same meaning and phonology as Hangul is processed in daily life, the activation threshold of its Hangul counterpart can be lowered. Even though we do not know whether the units of phonology or meaning are strengthened due to repetition effects, the two types of orthography appear to interact.

3. Experiment 2

Experiment 2 examined different word length effects for an alphabetic syllabary, Korean. Here, the focus is on the main orthography, Hangul. Korean Hangul has several characteristics that distinguish it from English, which might cause difference in word length effects. First of all, Korean has an almost perfect relationship

between letter and sound. In other words, Korean is a shallow orthography, whereas English is a deep orthography, having more complex relationships between letters and sounds. Thus, this characteristic can reduce word length effects for Korean because processing letters with inconsistent spelling sound rules in English would be more detrimental in processing a longer word. Second, each syllable of Korean is written with clear visual separation. A certain amount of space is put between every syllable in Korean. Thus, the physical separation between syllables might facilitate processing a multisyllabic word in Korean, whereas English writing does not confer any such advantage. If this is the case, processing of the multisyllable words should be easier for Korean words.

3.1 Method

Participant Forty native Korean graduate students and their wives. The students were enrolled at the University of Texas at Austin.

Materials and procedure Seventy-five words were selected from the Korean word frequency count. Three conditions were used: one syllable, two syllables, and three syllables. 25 words were selected for each condition. They were yoked in frequency across conditions according to the Korean word frequency count. The initial phoneme of each word was also matched across the three conditions. The task was to name a word on a computer screen after viewing 8 practice items. Word naming was performed as stimuli appeared on the screen. The experimenter checked for the pronunciation errors. Stimulus presentation and performance measurement were performed by 'Context' program. The latency and error rates for each condition were the dependant variables.

Statistical analysis The same methods to Experiment 1 were applied, except the independent variable (the word length)

3.2 Results and Discussion

Data containing mispronunciation and delay errors

were discarded. One participant whose error rate was over 10 percent was not included in data analysis. However, the overall error rates for each condition excluding that participant were less than one percent, which is meaningless in comparison. The median latency for each condition was computed for each participant. The means for these medians are presented in Table 3.

Table 3. Mean Naming Latencies, Standard Deviations, and Error rates

Syllable	RT	SD	Errors
One	560	84.7	0.2
Two	549	93.4	0.5
Three	571	91.9	0.7

Analysis of variance revealed that reading times were different across word conditions significantly ($F(2, 78) = 7.31, p < .01$) by subject. The analysis by item showed no statistical difference among conditions ($F(2, 48) = 0.95, p = .395$). However, the pattern of latency was V shape, indicating that latency did not increase monotonically across syllable length conditions. Specifically, two syllable words were processed faster than words of 1 or 3 syllables. Interestingly, one syllable words were processed more slowly than two syllable words.

The results of Experiment 2 demonstrated that the processing of two-syllable Korean words is faster than that of one-syllable words. Studies that investigate the processing of other orthographies including English words did not show no such pattern in any studies.

There could be the main reason on why the length effect is absent for Korean. One possible reason is that one syllable word is not so informative to bring any lexical activation since there are too many possible meanings in Korean one syllable words. In other words, there would be more cues to activate lexicon for multisyllable words than monosyllable words. Faster processing of two syllable words than one syllable words indicates that two syllable words contain more information on meaning used for word recognition. In

contrast, three syllable words took the longest, indicating the common processing difficulty from perceptual limitation when processing a long word with wider eccentricity. It overshadows the advantage of information increase due to the number of syllables. The boundary for this advantage would be two syllable.

4. General Discussion

This study was conducted in order to investigate the effect of two main lexical variables, the frequency and the length. Experiments one and two showed that word frequency effects and word length effects in Korean are quite different from English. Experiment 1 showed that the word frequency effect in Korean is not the same as those documented for English words [1,5,10]. Korean Hangul words with Hanja counterparts did not show word frequency effects that English words show. Experiment 2 demonstrated that two syllable words were named faster than one syllable words, which contradicted previous findings from other languages [12]. V shape word length effect might be one of the unique features in Korean Hangul. In addition, the two basic lexical effects in Korean orthographies differ from other orthographies.

The unreliable frequency effect in Korean Hangul recognition would be due to the usage of another orthography, Korean Hanja. Some ranges of frequency use Hangul and Hanja together, while other ranges do not. This indicates that any future study on Korean should take into consideration in selecting stimuli for experiment to avoid mixing two subtypes of Korean words in an experiment. In daily life, Hanja counterparts might be used more frequently for low frequency words to clarify the meaning of the word, which can provide advantage in processing that kind of words. Future studies might need to investigate the frequency effects for relatively low frequency words with Hanja counterparts, where the Hanja word

database must be reflected.

Alternatively, another lexical variable, the neighborhood size might have contributed to the null frequency effect. This study did not control the neighborhood size of a word across the condition although the neighborhood size has been proved the significant lexical variables that can affect the reaction time. More elaborate studies are needed in a subsequent study.

The absence of a typical word length effect indicates that less information is present in monosyllable words than in multisyllabic words. Specifically, the monosyllabic word is ambiguous in meaning because many semantic meanings are related with the one syllable as compared to two syllable words. For example, the monosyllable word “국” has many alternative meaning as compared to the disyllable word “국민.” Another possible processing that might have occurred is that the participants would think some monosyllabic words are not words. This possibility, however, can be dismissed because the error rates across all length conditions are not quite different.

Findings of the present study showed somewhat different results compared to the previously suggested word frequency and word length effects from other languages. This could be very informative because it might demonstrate some unique features of Korean alphabet system. In addition, difference in terms of processing Hangul words with and without Hanja counterparts will give us a hint about how to develop Korean language learning program.

It would be meaningful to test the effect of visual syllabic separation in the future. Would an artificial syllabic separation (e.g., fa ther) produce a diminished word length effect? It is possible that an separation would show slow reaction time overall, but patterns of reaction time between words with syllabic separations might differ substantially compared to words that do not have such separations.

References

- [1] Allen, P. A., McNeal, M., & Kvak, D. 1992, "Perhaps the lexicon is coded as a function of word frequency". *Journal of Memory and Language*, vol. 31, pp. 826-844. DOI: [http://dx.doi.org/10.1016/0749-596X\(92\)90041-U](http://dx.doi.org/10.1016/0749-596X(92)90041-U)
- [2] 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*, vol. 10, pp. 340-357. DOI: <http://dx.doi.org/10.1037/0096-1523.10.3.340>
- [3] Hauk, O., Davis, M.H., Ford, M., Pulvermuller, F., & Marslen-Wilson. 2005, "The time course of visual word recognition as revealed by linear regression analysis of ERP data". *Neuroimage*, vol. 30, pp. 1383-1400. DOI: <http://dx.doi.org/10.1016/j.neuroimage.2005.11.048>
- [4] Ellis, A.W., 2004. "Length, formats, neighbours, hemispheres, and the processing of words presented laterally or at fixation". *Brain and Language*. vol. 88, pp. 355-366. DOI: [http://dx.doi.org/10.1016/S0093-934X\(03\)00166-4](http://dx.doi.org/10.1016/S0093-934X(03)00166-4)
- [5] Forster, K. I., & Davis, C. 1984, "Repetition priming and frequency attenuation in lexical access". *Journal of Experimental Psychology: Learning, Memory and Cognition*, vol.10, pp. 680-698. DOI: <http://dx.doi.org/10.1037/0278-7393.10.4.680>
- [6] Fiebach, C.J., Friederici, A.D., Muller, K., von Cramon, D.Y., 2002. "fMRI evidence for dual routes to the mental lexicon in visual word recognition". *Journal of Cognitive Neuroscience*. vo. 14, pp. 11 - 23. DOI: <http://dx.doi.org/10.1162/089892902317205285>
- [7] Hudson, P. T. & Bergman, M. W. 1985, "Lexical knowledge in word recognition: word length and word frequency in naming and lexical decision tasks". *Journal of Memory and Language*, vol.24, pp. 46-58. DOI: [http://dx.doi.org/10.1016/0749-596X\(85\)90015-4](http://dx.doi.org/10.1016/0749-596X(85)90015-4)
- [8] Stockall, L., Stringfellow, A., & Marantz, A., 2004. "The precise time course of lexical activation: MEG measurements of the effects of frequency, probability, and density in lexical decision". *Brain and Language*. vol. 90, pp. 88 - 94. DOI: [http://dx.doi.org/10.1016/S0093-934X\(03\)00422-X](http://dx.doi.org/10.1016/S0093-934X(03)00422-X)
- [9] New, B., Rerrand, L., Pallier, C., & Brysbaert, M., 2006. "Reexamining the word length effect in visual word recognition: New evidence from the English lexicon project". *Psychonomic Bulletin & Review*. vol. 13, pp. 45-52. DOI: <http://dx.doi.org/10.3758/BF03193811>
- [10] Versace, R., & Nevers, B., 2003. "Word frequency effect on repetition priming as a function of prime duration and delay between the prime and the target". *British Journal of Psychology*. vol. 93, pp. 389-408. DOI: <http://dx.doi.org/10.1348/000712603767876299>
- [11] 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*", vol.118, pp. 43-71. DOI: <http://dx.doi.org/10.1037/0096-3445.118.1.43>
- [12] The Whaley, C. P. 1978, "Word-Nonword classification time", *Journal of Verbal learning and Verbal behavior*, vol.17, pp. 143-154. DOI: [http://dx.doi.org/10.1016/S0022-5371\(78\)90110-X](http://dx.doi.org/10.1016/S0022-5371(78)90110-X)
- [13] Coltheart, M. 1978, "Lexical access in simple reading tasks". In G. Underwood (ed.), pp. 151-216. San Diego, CA: Academic Press.
- [14] Park, K., & Lim, H. 2014, "A computational model explaining language phenomena on Korean visual word recognition". *Cognitive Systems Research*. vol. 27, pp. 11-24. DOI: <http://dx.doi.org/10.1016/j.cogsys.2013.04.001>
- [15] Yi, K. 1993, "On the role of frequency and internal structure in the processing of Kulja". *Korean Journal of Experimental and Cognitive Psychology*, vol.5, pp. 26-39.
- [16] Park, K. 1993, "Mental code involved in Hangeul word recognition". *Korean Journal of Experimental and Cognitive Psychology*, vol.5, pp. 40-55.
- [17] Simpson, G. B., & Kang, H. 1994, "The flexible use of phonological information in word recognition". *Journal of Memory and Language*, vol.33, pp. 319-331. DOI: <http://dx.doi.org/10.1006/jmla.1994.1015>
- [18] Lee, S. S., Lee, K. D., Nam, K. S., Chung, C. S., Lee, I. K., & Choi, Y. C. 1991, "Development of a corpus of contemporary Korean for compiling a lexicographical data base". Seoul, Korea: Korean Dictionary compilation board of Yonsei University.
- [19] Hulme, C., Neath, I., Stuart, G., Shostak, L., Surprenant, A. M., & Brown, G. 2006. "The distinctiveness of the word-length effect". *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 32, pp. 586-594. DOI: <http://dx.doi.org/10.1037/0278-7393.32.3.586>
- [20] New, B., Ferrand, L., Pallier, C., & Brysbaert, M. 2006. "Reexamining the word length effect in visual word recognition: New evidence from the English Lexicon Project". vol. 13, pp. 45-52.
- [21] Patberg, J. P., Dewitz, & Samuels, J. S. 1981, "The effect of context on the size of the perceptual unit used in word recognition". *Journal of Reading Behavior*, vol.13, pp. 33-48. DOI: <http://dx.doi.org/10.1080/10862968109547392>
- [22] Samuels, S. J., Laberge, D., & Bremer, C. D. 1978, "Unit of word recognition: evidence for developmental changes". *Journal of Verbal learning and Verbal behavior*, vol.17, pp. 715-720. DOI: [http://dx.doi.org/10.1016/S0022-5371\(78\)90433-4](http://dx.doi.org/10.1016/S0022-5371(78)90433-4)

Changhwan Lee

[Regular member]



- Aug. 1999 : Univ. of Texas, Austin, PhD in Psychology
- Sep. 2007 ~ present : Sogang Univ., Dept. of Psychology, Professor

<Research Interests>

Psychology of Reading, Cognitive Psychology

Yoonhyoung Lee

[Regular member]



- May. 2007 : Univ. of North Carolina a Chapel Hil, PhD n Psychology
- Sep. 2008 ~ Aug. 2013 : Catholic Univ. of Daegu, Dept. of Psychology, Professor
- Sep. 2013 ~ present : Yeungnam Univ., Dept. of Psychology, Professor

<Research Interests>

Cognitive Psychology, Cognitive Neuroscience

Tae Hoon Kim

[Regular member]



- Dec. 2009 : Ohio State Univ., PhD in Psychology
- Jan. 2010 ~ Feb. 2012 : Post Doc. in Ohio State Univ.
- Mar. 2012 ~ present : Kyungnam Univ., Dept. of Psychology, Professor

<Research Interests>

Cognitive Psychology, Motor Control