한국심리학회지 : 인지 및 생물
The Korean Journal of Cognitive and Biological Psychology
2023, Vol. 35, No. 4, 303-335
http://dx.doi.org/10.22172/cogbio.2023.35.4.004

Original Article

OPEN ACCESS

Syllable frequency effect in visual word recognition: a regression study on morphologically simple and complex Korean words

Joonwoo Kim[1], Solbin Lee[1], Sangyub Kim[2], Kichun Nam[1,2†]

[1]School of Psychology, Korea University
[2]Wisdom Science Center, Korea University

Previous studies on syllable frequency effect on visual recognition of Korean words have reported mixed results depending on their part of speech, morphological form (simple vs. inflected form), and measures of syllable frequency (e.g., type or token frequency). The present study reanalyzed lexical decision data from 23,173 words in three datasets (Yi et al., 2017; Kim et al., 2018; Kim et al., 2020). In eight independent hierarchical regression models, four frequency measures were employed for the orthographic and phonological first syllable, respectively. The first syllable frequency was measured in terms of type or token, or whether the length of syllabic neighbors was considered or not. The results showed robust facilitative effects of first syllable frequency in all forms of verbs, adjectives, and adverbs, while only in inflected form for nouns. The facilitative effects of first syllable frequency were demonstrated regardless of frequency measure, both in orthographic and phonological syllable frequency. This facilitative effect of syllable frequency could be due to the Korean writing system, which provides visually distinct syllabic boundaries, resulting in direct access to the mental lexicon via the orthographic route. The null effect of the syllable frequency in noun words could be attributed to their greater number of higher frequency syllabic neighbors (HFSN), compared to the other three categories. Thus, these results suggest that both the orthographic and phonological representation of a syllable plays an important role in visual word recognition.

Keywords: visual word recognition, syllable frequency effect, hierarchical regression, orthographic representation, phonological representation

1차원고접수: 23.01.02; 수정본접수: 23.04.16; 최종게재결정: 23.09.15

Copyright: © 2023 The Korean Society for Cognitive and Biological Psychology. This is and Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0(https://creativecommons.org/licenses/by-nc/4.0/) which permits use, distribution and reproduction in any medium, provided that the article is properly cited and the use is non-commercial.

어휘를 재인하는 과정에는 단어 빈도, 길이와 같은 어휘 요인과 음절, 형태소의 빈도 등 하위 어휘 요인이 중요한 역할을 담당한다. 그 중, 하위 어휘 단위에 해당하는 첫음절은 어휘의 가장 왼쪽에 위치하여 시각 어휘 재인 시에 가장 먼저 제공되는 정보로 다른 위치의 음절보다도 중요한 역할을 담당한다. 기존 연구에서 첫음절의 출현 빈도가 높을수록 시각 단어 재인 속도가 느려지고 정답률이 낮아지는 억제 효과가 보고된 바 있으며(Carreiras et al., 1993; Perea & Carreiras, 1998), 이처럼 첫음절을 공유하는 어휘들의 누적 빈도가 높은 경우에 어휘 재인 수행이 억제되는 현상을 음절 빈도 효과라 명명하였다(Carreiras et al., 1993). 음절 빈도 효과는 스페인어(Carreiras et al., 1993; Conrad et al., 2008), 프랑스어(Conrad et al., 2007; Grainger & Jacobs, 1996), 독일어(Conrad & Jacobs, 2004; Hutzler et al.,

† 교신저자: 남기춘, 고려대학교 심리학부, (02841) 서울 성북구 안암로 145 법학관 구관 405호, E-mail: kichun@korea.ac.kr

2004)와 같은 인도유럽어 계통의 언어뿐만 아니라, 비인도유럽어인 한국어에서도 영향을 미치는 것으로 나타났다(Kwon & Nam, 2011; Kwon, 2012). 그러나 음절 빈도 효과의 방향성은 한국어 시각 어휘 재인에서 억제적(e.g., Kwon, 2012) 혹은 촉진적 효과가 보고되거나(e.g., Kim & Nam, 2018), 효과가 없는 것으로 보고되는 등(e.g., Jin et al., 2018), 다소 혼재된 결과가 나타났다. 본 연구는 기존 한국어 연구에서 혼재된 음절 빈도 효과에 영향을 미치는 요인에 대한 개관적 설명 이후, 이러한 요인을 고려하여 기존의 대규모 시각 어휘 재인 데이터를 대상으로 일관된 기준을 사용한 재분석을 통해 한국어 시각 어휘 재인에서 음절의 역할을 확인하고 음절 빈도 효과의 방향성을 조사하고자 한다.

한국어 시각 어휘 재인에서 음절 빈도 효과에 영향을 미치는 요인 중 첫 번째는 음절 빈도의 계산 방식이다. 먼저 음절 이웃(syllable neighborhood)은 목표 단어의 첫 번째 음절을 공유하는 단어의 집합으로 정의되며, 음절 빈도(syllable frequency)는 동일한 첫음절을 공유하는 이웃 단어의 가짓수를 의미하는 타입 빈도(type frequency), 모든 이웃 어휘의 출현 빈도를 합산한 토큰 빈도(token frequency)로 분류할 수 있다(Conrad et al., 2008). 스페인어 단어를 대상으로 이루어진 초기 음절 빈도 효과 연구에서는 타입·토큰 빈도 간의 구분이 이루어지지 않아 고빈도 첫음절 단어에 대한 어휘 판단 과제 수행이 저빈도 첫음절 단어 대비 저하되는 억제적 효과를 음절 빈도 효과로 설명하였다(Carreiras et al., 1993). 그러나 이후 같은 스페인어 대상 연구에서 첫음절의 타입·토큰 빈도 간 상이한 효과를 보고함에 따라, 음절의 타입 및 토큰 빈도 간의 구별의 필요성이 제기되었다(Conrad et al., 2008). 해당 연구에서는 음절의 타입 빈도 증가에 따라 어휘 판단 시간이 감소하는 촉진 효과가 나타난 반면, 토큰 빈도 증가는 어휘 판단 시간이 증가하는 억제 효과가 나타났다. 저자들은 이에 대하여 타입 빈도는 전형성(typicality)를 반영하는 변인이며, 따라서 타입 빈도에 의한 촉진적 효과는 어휘 전단계(prelexical stage)에서의 전반적 어휘 활성화 수준 증가로 인한 빠른 추측의 가능성 증가로 설명하였다. 반면 토큰 빈도는 심성 어휘집의 어휘 접속(lexical access) 단계에 직접 개입하는 변인으로서, 토큰 빈도의 억제적 효과는 빈도가 높은 이웃 단어의 존재로 인하여 외측 억제(lateral inhibition)가 발생하고, 따라서 경쟁 후보군 표상에서 목표 단어를 재인하는 데 어려움을 겪기 때문으로 설명하였다. 이와 유사하게 한국어 어휘 판단에서 타입 및 토큰 빈도의 효과를 비교한 Kwon(2012)은 단계적 회귀분석을 통해 기존 한국어 명사 단어 대상 어휘 판단 연구(Kwon & Nam, 2011; Lee & Kwon, 2012; Lee & Taft, 2010)에 사용된 자극의 음절 빈도를 타입 및 토큰 빈도 계산 방식으로 구별하여 재분석을 시도하였다. 그 결과, 음절 토큰 빈도의 억제 효과가 스페인어 연구(Conrad et al., 2008)와 동일하게 나타났으나, 음절 타입 빈도는 오직 유사 단어에서만 유의한 억제 효과를 보였다. 그러나 최근의 한국어 시각 어휘 재인 연구에서 음절 타입 빈도의 효과뿐 아니라(Kwon, 2012; Kwon & Lee, 2015; Jin et al., 2018), 타입 빈도 역시 일관된 효과를 보고하지 않았다(Kwon & Lee, 2015; Jin et al., 2018). 따라서 한국어 시각 어휘 재인 시 나타나는 음절 빈도의 효과를 조사하기 위해 음절 빈도에 대한 정의를 구별하여 타입 및 토큰 빈도라는 상이한 계산 방식에 따른 효과를 비교할 필요성이 제기된다.

음절 빈도 효과의 혼재된 결과에 영향을 미치는 두 번째 요인은 음절의 철자, 혹은 음운 표상에 따른 연구 간 차이이다. 음절 빈도 효과를 보고한 기존 연구들(Carreiras et al., 1993; Hutzler et al., 2004; Conrad et al., 2007; Kwon, 2012)은 음절이 음운 정보를 표상하는 하위 어휘 요인에 해당한다고 가정하였다. 예를 들어, Carrieras 등(1993)은 음절이 두 철자(bigram)와는 구별되는 음운형 단위에 해당한다고 제안하였으며, Conrad 등(2007)에서는 프랑스어 시각 재인에서 음절이 음운형으로 정의되었을 때 유의미한 억제 효과를 보고하였다. 한국어에서도 음절의 음운 정보에 의하여 시각 단어 재인이 억제된다는 결과가 나타나(Kwon & Nam, 2011), Conrad와 동료들의 연구(2008)에서 제안한 바와 같이 음절이 표상하는 정보는 음운 정보에 해당하며, 음운 경로를 따라 심성 어휘집으로의 접속이 이루어진다고 볼 수 있다. 그러나 인도-유럽어 계통의 언어와 달리, 한국어에서는 시각 단어 재인에서 음절이 음운 정보가 아닌 철자 정보를 우선적으로 표상한다는 주장이 제기되었다(Bae & Yi, 2010; Koo et al., 2012; Choi et al., 2015). Bae와 Yi(2010)는 형태 점화 어휘 판단 과제(form primed lexical decision task)를 통하여, 점화 단어와 목표 단어 간의 철자만이 일치하는 조건(논란-*논의*), 음운만이 일치하는 조건(논란-*놀이*), 음운-철자 모두 일치하는 조건(논란-*논리*), 무관련 조건(논란-*조미*)의 어휘 판단 반응시간을 비교하였다. 실험 결과, 철자만 일치하는 조건과 음운-철자가 모두 일치하는 조건은 무관련 조건 대비 유의미한 촉진 효과가 나타난 반면, 음운만 일치하는 조건은 무관련 조건 대비 유의미한 억제 효과가 나타났으며, 이와 같은 결과는 최근의 Choi 등(2015)이나 Lim 등(2022)의 연구에서도 보고된 바 있다. 이를 토대로 저자들(Bae & Yi, 2010)은 심성 어휘집으로의 어휘 접속(lexical

access) 시에 반드시 음운 경로를 경유해야 한다는 기존 연구(e.g., Conrad et al., 2007; Kwon, 2012)의 주장을 반박하였으며, 한국어 시각 재인에서 음절이 우선적으로 표상하는 정보가 철자 정보라고 논하여, 철자 음절에서 심성 어휘집으로 직접 연결되는 시각 단어 재인 모형을 제안하였다. 이처럼 한국어 시각 재인에서 음절의 표상 정보가 철자 정보인지 음운 정보인지에 대한 논의가 지속적으로 이루어져 왔으나, 단일한 연구에서 철자형, 음운형 음절을 구분하고 각각의 음절 빈도가 보이는 효과를 조사한 결과는 보고된 바가 없다. 이에 본 연구는 동일한 음절의 철자형 및 음운형 빈도를 구분하여 산출하고 그 효과를 비교 분석하여, 한국어 시각 어휘 재인에서 음절의 음운 및 철자 표상이 수행하는 역할을 확인하고자 한다.

다음으로, 어휘의 음절 이웃의 길이 역시 한국어 시각 재인에서 음절 빈도 효과의 혼재와 연관이 있는 요인으로 볼 수 있다. 초기 Coltheart와 동료들에 의해 제안된 표기 이웃(orthographic neighbor)의 개념은 목표 단어에서 단일한 낱자(letter)를 대체하고, 그 외 모든 낱자는 위치를 동일하게 유지하는 단어들의 집합을 의미하며, 이렇게 형성된 단어의 이웃 크기가 Coltheart's *N*으로 정의된다(Coltheart et al., 1977). 예를 들어, 영어 단어 "cat"의 표기 이웃은 "bat", "hat", "mat"과 같은 3개 철자 길이의 단어가 해당한다. 그러나, 한국어, 스페인어, 프랑스어 등의 언어는 얕은 철자 체계(shallow orthograpahy)를 가져 음절 간의 경계가 명확하고, 따라서 음절 단위가 이웃의 단위로 사용될 수 있다. 이러한 언어들에서 음절 이웃(syllabic neighborhood)은 기존의 단어 이웃(i.e., Coltheart's *N*)과 달리, 목표 단어의 첫번째 음절을 공유하는 단어들의 집합으로 정의된다. 예를 들어, '고구마'라는 단어의 이웃은 '고데기', '고도', '고리타분'과 같은 단어들이 해당된다. 이와 같이, 음절을 기반으로 한 이웃 단어와 낱자를 한 이웃 단어는 어휘 길이라는 측면에서 차이를 보이고 있는데, 음절 기반 이웃 단어 중에서도 고빈도 이웃 단어에 의해 시각 단어 재인이 억제된다는 결과가 보고된 바 있어(Perea & Carreiras, 1998; Conrad et al., 2008), 이웃 단어의 길이에 기반한 음절 빈도의 차이를 연구할 필요성이 제기된다. 더욱이, 기존 음절 빈도 효과를 보고한 해외 및 한국어 명사 단어 연구들은 공통적으로 2음절 명사 단어를 실험 자극으로 사용하여, 다음절 단어의 시각 재인에서 음절이 미치는 영향력을 탐구하였다(Carreiras et al., 1993; Perea & Carreiras, 1998; Conrad et al., 2008; Kwon, 2012; Kwon & Lee, 2015; Jin et al., 2018). 그러나, 최근의 한국어 연구에서는 2-8음절의 다양한 길이를 지닌 자극을 사용하여 시각 어휘 재인에서 음절이 미치는 영향력을 연구하였다(Kim & Nam, 2018; Kim et al., 2020; Kwon et al., 2023). 이러한 단어 및 어절 길이의 차이는 심성 어휘집 상에서 이웃 후보군이 활성화될 때, 동일한 어휘 길이를 지닌 후보군만 활성화되는지, 아니면 첫음절을 공유하는 모든 길이의 이웃 후보군이 활성화되는지에 대한 의문을 제기한다. 그러나 어휘 길이에 따른 음절 빈도의 차이를 세밀하게 구분하여 조사한 연구는 아직 수행된 바가 없기에, 본 연구에서는 어휘 길이를 제한하여 산출한 음절 빈도와 길이를 제한하지 않고 산출한 음절 빈도의 영향력 차이를 비교 분석하고자 한다.

어휘의 음절 수는 음절 빈도에 대한 조작적 정의가 상이하게 이루어지도록 한 요인 중 하나이다. 단어를 대상으로 한 연구(e.g., Álvarez et al., 2004; Kwon & Nam, 2011; Kwon, 2012)는 모두 음절 이웃에 대하여 전통적인 이웃의 개념, 즉 동일한 음절로 시작하는, 음절 수가 동일한 단어를 적용하여 빈도를 계산하였으나, 최근의 어절을 대상으로 한 연구(e.g., Kim & Nam, 2018; Kim et al., 2020)에서는 음절 수를 제한하지 않고 동일한 음절로 시작하는 어휘 항목의 집합을 음절 이웃으로 상정하여 음절 빈도를 계산하였다. 따라서, 한국어 어휘 재인에서 나타나는 상이한 음절 빈도 효과는 어휘의 형태뿐 아니라, 음절 수 제한 여부에 따른 차이에 기인하였을 가능성도 배제하기 어렵다. 이와 관련하여 Chetail과 Mathey(2011)는 기존에 프랑스어에서 보고된 억제적 음절 이웃의 효과가 음절 이웃의 계산 방식, 즉 음절 수를 기준으로 한 이웃의 길이에 따라 상이하게 나타나는지 여부를 조사하였다. 그 결과, 고빈도 음절 이웃 수, 즉 해당 단어의 출현 빈도보다 높은 빈도를 가진 음절 이웃의 수(number of higher frequency syllabic neighbors; HFSN)의 억제적 효과는 동일한 음절 수(2음절)를 가진 이웃들을 기준으로 계산되었을 때에만 유의하게 나타났으며, 2음절 이상의 임의의 음절 수를 가진 다음절 단어를 이웃으로 가정하여 계산된 경우 그 효과가 유의하지 않았다. 따라서, 어휘 항목의 음절 수를 제한한 전통적 음절 이웃의 정의를 기반으로 한 음절 빈도와 음절 수를 제한하지 않은 음절 이웃의 정의를 기반으로 한 음절 빈도의 효과에 대한 비교 또한 필요하다고 볼 수 있다.

한편, 이웃 중 해당 단어의 출현 빈도보다 높은 빈도를 가진 어휘를 의미하는 고빈도 이웃(higher frequency neighbor) 역시 어휘 재인에 주요한 영향을 미치는 변인으로 고려되기도 하였다. 예를 들어, Grainger와 Jacobs(1996)는 프랑스어 어휘 재인 과정에 미치는 이웃 단어의 영향을 조사

하고자 어휘 판단 과제를 실시하였으며, 그 결과 단어의 이웃 크기[1]는 촉진 효과가 나타난 반면, 이웃 크기가 일정하게 통제된 경우 고빈도 이웃이 존재할 경우 억제 효과가 나타났다. 그러나 고빈도 이웃의 수는 토큰 빈도가 증가할 경우 함께 증가하는 정적 상관의 경향을 보이며, 따라서 토큰 빈도와 고빈도 이웃 수는 혼입의 여지가 존재한다(Perea & Carreiras, 1998). 예를 들어, Perea와 Carreiras(1998)는 어휘 재인 및 산출 과정에서 음절 빈도의 효과를 조사하기 위해 어휘 판단 과제와 명명 과제를 실시하였다. 그 결과, 어휘 판단 과제(실험 1)에서 첫음절 토큰 빈도의 억제 효과가 유의하게 나타난 반면, 명명 과제(실험 2)에서는 첫음절 토큰 빈도의 촉진 효과가 나타났다. 그러나 실험 1, 2에 대한 사후 분석 결과, 어휘 판단 과제에서 음절의 토큰 빈도는 유의하지 않게 나타난 반면, 고빈도 음절 이웃 수(HFSN)는 유의한 억제 효과를 보였다. 또한, 고빈도 음절 이웃 수를 직접적으로 조작하여 실시한 어휘 판단 과제(실험 3)에서 역시 고빈도 음절 이웃 수의 유의한 억제 효과를 보였다. 따라서, 어휘 재인 과정에 미치는 음절 빈도의 영향을 조사하기 위해서는 고빈도 음절 이웃에 대한 적절한 통제가 필요하다고 할 수 있다. 실제로 앞서 소개한 Conrad 등(2008)은 실험 1에서 나타난 음절 타입 및 토큰 빈도의 상이한 효과가 고빈도 음절 이웃 수의 혼입에 의해 나타났을 가능성을 배제하기 위하여, 실험 2에서는 이를 통제한 뒤의 음절의 타입, 토큰 빈도 효과를 조사하였으며, 실험 1과 동일하게 두 빈도 간의 상이한 패턴을 관찰한 바 있다.

이와 더불어, 음절이 가진 형태론적 요인(morphological factor)이 음절 빈도 효과에 영향을 준다는 결과 또한 보고된 바 있다(Kwon & Nam, 2011; Kwon et al., 2012). Kwon과 Nam (2011)은 음운 및 철자 이웃 효과에 미치는 한자어 형태소 요인의 영향을 조사하고자 두 개의 어휘 판단 실험을 통해 첫음절의 한자어 이웃 크기[2]에 따른 음운 이웃과 철자 이웃 크기의 효과를 비교하였다. 그 결과, 실험 2에서 철자 이웃 크기의 효과는 유의하지 않았으나, 실험 1에서 음운 이웃 크기 효과는 억제적으로 나타났으며 그 효과 크기가 한자어 형태소 이웃 크기가 큰 조건에서 감소하는 상호작용이 나타났다. 결론적으로, 이전의 연구 결과는 적어도 한국어 명사 단어 재인 과정에서의 음절은 철자론적 측면, 음운론적 측면 그리고 마지막으로 형태론적 측면으로 구분되어 작용한다고 볼 수 있으며, 철자 및 음운 음절 빈도 효과의 비교를 위해서는 음절 자체가 가진 형태론적 측면에 대한 고려가 필요하다고 할 수 있다.

그러나 실제 한국어 문장에서 단어는 단독으로 사용되기보다 주로 다양한 기능 형태소(문법 형태소)가 결합된 형태, 즉 어절이 사용된다. 예를 들어, “선생님께서 학교에 가셨다”와 같은 문장에서, 술어인 “가시었다”(가- + -시- + -었- + -다)는 동사 어간인 “가-”에 높임말을 가리키는 선어말어미 “-시-”와 과거 시제를 나타내는 선어말어미 “-었-”, 마지막으로 종결 어미인 “-다”가 결합된 형태인 용언 어절이다. 어절은 용언 활용뿐 아니라 체언에서도 사용되는데, 앞선 예시 문장의 주어 “선생님께서”(선생님 + -께서)와 부사어 “학교에”(학교 + -에)는 모두 명사 어근에 조사(각각 주격 조사 “께서”, 부사격 조사 “-에”)가 결합된 형태인 체언 어절이다. 이러한 측면에서 최근에는 한국어 단어가 아닌 어절 재인 과정에 대한 연구 또한 활발하게 이루어지고 있다(Kim & Nam, 2018; Kim et al., 2020; Kim et al., 2022). 이러한 어절 대상 연구와 기존 단어 대상 연구에서 첫음절의 효과가 다르게 나타났는데, 주로 명사 단어를 사용한 연구에서는 첫음절의 억제 효과가 나타난 반면(e.g., Kwon & Nam, 2011; Kwon, 2012), 어절 대상 연구에서는 첫음절의 촉진 효과가 보고되었다(e.g., Kim & Nam, 2018; Kim et al., 2020; Lee et al., 2019). Kim과 Nam(2018)은 한국어 체언 어절과 용언 어절에 대한 어휘 판단 시간에 대하여 어휘 요인과 하위 어휘 요인을 투입하여 다중 회귀 분석을 실시하였고, 분석 결과 첫음절의 빈도가 높을수록 반응시간이 빨라지는 것으로 나타났다. 체언 어절만을 사용한 어휘 판단 실험에 대한 다중 회귀 분석 연구(Kim et al., 2020)와 용언 어절만을 사용한 어휘 판단 실험 연구(Lee et al., 2019) 모두에서 앞선 결과와 같은 촉진 효과가 보고된 바 있어, 한국어 어절과 단어 재인 과정에서 하위 어휘 단위인 음절이 미치는 효과와 그 처리 방식의 차이에 대한 논의의 필요성이 대두되고 있다. 이에 본 연구에서는 기존 한국어 단어와 어절을 기반으로 진행된 어휘 판단 데이터에 대한 상관 다중 회귀 분석을 통하여 음절 빈도 효과를 재검증하고, 시각 어휘 재인 과정에서 음절 단위가 갖는 의의를 탐구하고자 하였다.

기존의 대규모 어휘 판단 데이터 분석 연구에서는 주로 다중 회귀 분석(multiple regression)을 통하여 다양한 어휘

1) Grainger와 Jacobs(1996)에서 단어 이웃은 Coltheart 등(1977)이 정의한 Coltheart’s *N*, 즉 목표 단어와 낱자 하나를 제외한 나머지 모든 위치의 낱자가 동일한 단어의 집합으로 정의되었다.

2) 해당 연구에서 한자어 이웃은 목표 단어의 첫음절과 동일한 한자어로 시작되는 단어들의 집합으로 정의되었다. 예를 들어, 국가(國家)의 한자어 이웃은 “국민”, “국고”, “국회”, “국토”, “국어” 등으로 볼 수 있는 반면, 표기 이웃은 이외에도 “국화”, “국자”, “국수”, “국물” 등의 단어가 포함되며, 음운 이웃은 위 예시에서 첫음절이 “/궁/”으로 발음되는 “국민”, “국물”이 제외된 단어의 집합으로 볼 수 있다.

및 하위 어휘 요인이 어휘 재인, 혹은 명명 과정에 미치는 영향을 조사하였다(Baayen et al., 2006; Balota et al., 2004; Goh et al., 2016; Yap & Balota, 2009). 예를 들어, Balota 등은 영어 단음절 단어 2,428개 자극에 대한 어휘 판단 및 명명 데이터를 대상으로 분석을 시도하였으며(Balota et al., 2004), Yap과 Balota(2009)는 분석 대상을 다음절 단어로 확장하여 6,115개 자극의 어휘 판단 및 명명 데이터에 대한 분석을 진행하였다. 두 연구 모두에서 위계적 다중 회귀 분석 기법(hierarchical multiple regression technique)이 사용되었으며, 단계별 모형의 설명력(R 제곱; R squared)의 변화량을 비교하였다. 한편, Goh와 동료들의 연구(2016) 역시 말소리 단어 재인 과정에서 의미 요인의 효과를 조사하기 위하여, 참가자 80명을 대상으로 514개 명사 단어에 대한 어휘 판단 및 의미 분류 과제를 실시하고, 해당 대규모 데이터를 대상으로 위계적 다중 회귀 분석 기법을 사용하였다. 이와 더불어, 저자들은 어휘 길이, 단어 빈도 등을 포함한 어휘 변인과 구체성(concreteness), 정서가(valence) 및 각성가(arousal) 등을 포함한 의미 변인이 위계적 회귀 분석에서 보인 효과가 과제 유형(어휘 판단 과제 및 의미 분류 과제)에 따라 상이하게 나타나는지를 조사하고자 선형 혼합 효과 모형(linear mixed effect model)을 사용하여 각 변인과 과제 요인 간의 상호작용을 추가 분석하였다.

한국어에서도 이와 유사하게 어휘 판단 데이터에 미치는 다양한 어휘 요인 및 하위 어휘 요인이 미치는 영향을 위계적 다중 회귀 분석을 통하여 조사한 연구가 진행된 바 있다(Kim & Nam, 2018; Kwon, 2012). 한국어 시각 어휘 재인에서는 Kwon(2012)이 기존 어휘 판단 데이터(Kwon & Nam, 2011; Lee & Kwon, 2012; Lee & Taft, 2010)를 대상으로 첫음절 타입 및 토큰 빈도의 효과를 비교 분석하였으며, 세 연구의 명사 단어 114개가 분석 재료가 되었다. 그러나 저자가 논의에서 연구의 한계점으로 언급한 바와 같이 자극의 수가 국외 연구에 비해 상대적으로 적었으며 명사 자극만을 분석하였다는 점은 보다 규모가 큰 데이터를 대상으로 어휘의 다양한 품사 및 형태에 대한 추가 연구에 대한 필요성을 제기한다. 이에 본 연구에서는 기존에 명사 단어 20,959개와 명사 어절 480개, 용언 원형 어절 1,093개, 용언 굴절형 어절 406개를 포함하여 총 22,938개의 어휘를 대상으로 진행된 대규모 어휘 판단 데이터를 재분석하여 한국어 시각 어휘 재인 과정에 미치는 음절 빈도의 효과를 면밀히 재조사하고자 한다. 분석 방법으로는 위계적 다중 회귀 분석 기법을 사용하여 다른 변인들의 설명 변량을 제거한 후에 관심 변인의 잔여 변량에 대해 첫음절 빈도만의 설명 변량의 유의미성 여부를 조사하고자 하였다.

본 연구의 첫 번째 목적은 음절이 한국어 단어와 어절의 시각 재인에서 각기 어떻게 기능하는지를 조사하는 것이다. 어절의 정의는 어절에 포함된 어근 혹은 어간의 품사에 따라 달라질 수 있다. 예를 들어, 어휘 판단 메가스터디인 Yi 등(2017)에서는 명사의 경우 단어 형태를 사용하였으며, 동사와 형용사로 대표되는 용언의 경우 “먹다”(먹- + -다), “가엾다”(가엾- + -다)와 같이 사전에 등재되어 있는 원형이 사용되었다. 이는 명사 단어에 상응하는 용언의 단어 형태가 용언의 원형이라고 가정하였던 것으로 보인다. 단어의 정의를 국립국어원 표준국어대사전(1999; 2008)에 등재되어 있는지 여부에 의해 결정한다면 이들 용언 원형은 단어로 볼 수 있으며 국립국어원의 단어 정의에 따라서도 단어로 볼 수 있다[3]. 이처럼 용언의 원형은 홀로 언어 단위를 구분하는 측면에서 보면 단어로 볼 수 있고, 문장의 최소 구성 성분으로 보거나 사용되는 실태로 보면 어절로 구분할 수 있다. 따라서 본 연구에서는 용언의 원형 자료는 특별히 굴절형 어절과 구분하였으며, 체언의 경우에도 용언의 원형에 대응하는 단어와 용언의 굴절형에 대응하는 어절을 구분하여 자료를 분석하였다.

본 연구의 두 번째 목적은 음절 빈도 효과가 첫음절의 철자 정보에 의한 것인지, 혹은 음운 정보에 의한 것인지에 대해 탐구하는 것이다. 기존 단어 재인 연구에서는 주로 첫음절을 음운론적 단위(phonological unit)로 보았으며, 음운 이웃의 표상으로 인하여 첫음절의 억제 효과가 나타난다고 설명한다. 음절의 음운형 정보는 한국어 명사 단어 재인 연구에서도 나타났다. 그러나 시각 표상 과정에서 음절이 음운론적 단위라 할지라도 철자 정보를 배제하기 어렵다는 의견이 제기된 바 있다(Conrad et al. 2009). 단어가 아닌 어절을 사용한 한국어 연구에서는 주로 철자형 음절 빈도를 사용하여 첫음절 빈도의 촉진적 효과를 보고한 바 있으나, 음운 변동을 고려한 음운형 음절 빈도를 동시에 사용한 음절 빈도 효과는 밝혀진 바가 없다. 이에 본 연구에서는 단어와 어절을 포함한 한국어 어휘 재인에서 첫음절의 철자형 빈도뿐 아니라, 첫음절의 음운 변동을 고려한 음운형 빈도 또한 사용하여 재분석하고자 한다.

3) 국립국어원에서는 단어의 정의를 다음 3가지로 정의하고 있다. 1. 분리하여 자립적으로 쓸 수 있는 말. 2. 분리하여 자립적으로 쓸 수 있는 말에 준하는 말. 3. 앞의 1~2의 말 뒤에 붙어서 문법적 기능을 나타내는 말. 반면에 어절에 대해서는 국립국어원에 아래처럼 정의하고 있다. “어절은 문장을 구성하고 있는 각각의 마디이다. 문장 성분의 최소 단위로서 띄어쓰기의 단위가 된다.” 이러한 정의를 바탕으로 할 때 앞서 예시로 제시된 “먹다”와 같은 동사 원형은 어절이라고도 볼 수도 있다.

본 연구는 기존에 수집된 대규모의 어휘 판단 데이터에 대해 단어(원형) 및 어절, 체언과 용언으로 분류하여 상관 및 다중 회귀 분석을 시행하였다. 첫 번째로 고려된 변인은 어휘의 전체 형태 빈도와 음절 수 변인이었다. 다음으로, 음절 빈도 효과에 영향을 미치는 것으로 보고된 변인을 고려하고자, 첫음절의 의미 여부, 첫음절의 철자-음운의 일치 여부, 첫음절의 고빈도 이웃 수(higher frequency syllable neighbor, HFSN)를 분석에 포함하였다. 마지막으로, 현재까지 다양하게 정의되었던 첫음절 빈도를 어휘의 음절 수를 제한하여 계산한 빈도와 제한하지 않고 계산한 빈도, 타입과 토큰 빈도, 철자 음절과 음운 음절 등 총 8가지 방식으로 계산하였다. 기존 연구 결과로 비추어 볼 때, 본 연구의 예상 분석 결과는 다음과 같다. 한국어 어휘 재인 중, 단어 재인에서는 음운형 음절 빈도의 억제적 효과가 유의하게 나타날 것으로 예상되지만, 어절 재인에는 철자형 음절 빈도가 촉진적 효과가 나타날 것으로 예상된다.

방 법

참가자

Yi 등(2017)의 어휘 판단 과제 메가스터디(megastudy)에 참여한 참가자는 국내 소재 4개 종합대학에서 선발한 58명이었으며, 이후 분석 과정에서 6명이 제외되어 최종 52명의 데이터가 분석 대상에 포함되었다. Kim 등(2020)의 실험 1에는 36명이 참가하였으나, 성실하게 실험에 응하지 않은 참가자와 우세 손을 오른손으로 통제하기 위하여 왼손잡이, 양손잡이 참가자를 포함한 4명을 제외하여 최종 32명의 데이터가 분석에 포함되었다. Kim과 Nam(2018)에는 참가자 21명이 참가하였으며, 제외된 참가자 없이 모든 참가자의 자료가 분석에 포함되었다. 세 자료를 종합하여 총 105명의 어휘 판단 과제 수행 결과가 분석 대상이 되었다.

절차

실험 절차로는 세 자료 모두 시각 어휘 판단 과제를 사용하였으며, 참가자들은 화면에 제시되는 목표 자극이 한국어에서 사용되는 단어 또는 어절일 경우 “yes” 반응을, 그렇지 않을 경우 “no” 반응을 하도록 하였다. 첫 번째 자료인 Yi 등(2017)에서는 총 30,930개의 단어를 대상으로 어휘 판단 과제를 시행하였으며, 음절 수는 1~5개로, 품사는 명사, 동사, 부사, 형용사 등으로 국한되었다. 비단어도 단어와 동일한 개수가 생성되었으며, 비단어를 구성하는 방식은 단어에서 특정 자모 하나를 변형시키는 방식을 사용하였다. 실험 과제는 시각 어휘 판단 과제를 사용하였으며, 실험 절차는 다음과 같다. 화면 중앙에 응시점(+)이 200ms 동안 제시되었으며, 이후 공백 화면이 200ms 동안 제시된 다음에 단어 혹은 비단어가 제시되었다. 참가자들은 제시된 텍스트가 단어인지 여부를 판단하여, 맞으면 버튼 박스의 오른쪽 버튼을, 단어가 아니면 왼쪽 버튼을 누르도록 지시받았다. 모든 참가자들은 하루 최대 4시간까지 실험을 수행하였으며, 참가자당 실험 시간은 32시간 가량 소요되었다. 실험은 124개 블록으로 구성되었으며, 한 블록당 512개의 자극이 할당되었다. 참가자가 실험을 완료하는데 소요된 기간은 최소 3주, 최대 6개월이었다.

두 번째 자료에 해당하는 Kim과 Nam(2018)에서는 총 2개의 실험을 시행하였으며, 실험 1에서는 체언 어절만을, 실험 2에서는 용언 어절만을 대상으로 하여 시각 어휘 판단 과제를 시행하였다. 실험 1에서는 총 21명의 대학생들이 참여하였으며, 연령 범위는 20~29세였다. 실험 1의 시각 자극으로는 체언 어절 240개와 비어절 240개로 총 480개의 자극이 선정되었으며, 어절 자극은 세종 코퍼스의 비율을 참고하여 선정되었다(신문 20%, 영화 10%, 전문지식도서 30%, 인터넷 텍스트 40%). 실험 절차는 다음과 같다. 화면 중앙에 응시점(+)이 500ms 동안 제시되었으며, 이후 1000ms 동안 목표 자극이 제시되었다. 참가자들은 제시된 자극이 한국어에서 실제 쓰는 어절인지 여부를 판단하여, 맞으면 키보드의 ‘m’ 키를, 아니라면 ’z’ 키를 누르도록 지시받았다. 실험 2의 용언 어절 실험에서도 동일하게 21명의 대학생이 참여하였으며, 참가자들은 실험 1에 참여하지 않은 인원으로 구성되었다. 자극의 개수도 실험 1과 동일하게 어절 240개, 비어절 240개로 총 480개의 자극을 사용하였으며, 용언 어절 240개 중 절반은 형용사 어절, 나머지 절반은 동사 어절로 구성되었다. 실험 과제 및 진행 절차는 실험 1과 동일하게 진행되었다.

세 번째 자료인 Kim 등(2020)에서는 체언 어절만을 대상으로 실험을 진행하였으며, 실험 1에서는 자극을 중앙 시야에 제시하였으며, 실험 2에서는 자극을 좌시야 혹은 우시야에 제시하는 편시야 제시 방식을 채용한 어휘 판단 과제를 활용하였다. 본 연구에서는 Kim 등(2020)의 실험 중, 나머지 두 자료와 같이 중앙 시야에 자극이 제시된 실험 1에 대한 데이터만을 재분석하였으며, Kim 등(2020)의 실험 자극 및 절차는 다음과 같다. 체언 어절 300개와 그에 대응되는 비어절 300개로 총 600개의 어절이 사용되었으며, 사용된 체언 어절의 구성 비율은 신문 20%, 영화 10%, 전문지식도서 30%, 인터넷 텍스트 40%에 해당하였다. 선정된 자극들에

대한 빈도 추출은 세종 코퍼스(Kang & Kim, 2009)를 기반으로 진행되었다. 실험 과제는 어휘 판단 과제로 응시점(+)이 화면 중앙에 500ms 동안 제시되었으며, 그 후에 목표 어절 혹은 비어절이 180ms 동안 제시되었다. 텍스트가 제시된 후에 공백 화면이 2000ms 동안 제시되었으며, 공백 화면이 나타나는 동안, 참가자들은 제시된 자극이 어절이었는지 여부를 판단하여, 어절이면 오른손으로 키보드의 '/' 키를, 아니면 왼손으로 'z' 키를 누르도록 지시받았다.

재료

실험 재료로는 Yi 등(2017)의 한국어 명사 단어와 용언 원형 어절을 사용한 대규모 어휘 판단 과제 데이터와 한국어 명사 어절과 용언 굴절형 어절을 사용한 Kim과 Nam(2018), Kim 등(2020)의 어절 판단 과제 데이터를 사용하였다. Yi 등(2017)은 총 30,930개의 어휘를 사용하였으며, 그 중 명사 단어는 21,996개, 용언 원형 어절(용언 어간에 종결 어미 '-다'가 결합된 형태)은 부사 1,264개, 동사 6,226개, 형용사 1,744개 등을 포함하여 9,234개였다. Kim과 Nam(2018)은 명사 어절(명사 단어에 조사가 결합된 형태) 179개, 용언 굴절 어절(용언 어간에 종결 어미 '-다' 이외의 어미가 결합되었거나 종결 어미 '-다'와 함께 선어말 어미가 결합된 형태) 130개, 원형 어절 12개를 포함하여 총 429개의 어휘를 사용하였다. Kim 등(2020)은 명사 어절 300개와 용언 굴절 어절 294개, 용언 원형 어절 6개를 포함하여 총 600개 어휘를 사용하였다. 상관 및 회귀 분석에서 전체 형태 빈도(full-form frequency)를 고려하기 위하여, 1500만 어절 규모의 세종 말뭉치(Kang & Kim, 2009)에 포함된 어휘 형태만을 선별하여 분석에 포함하였다[4]. 그 결과, Yi 등(2017)에서는 반응시간 데이터가 누락된 15개 항목(명사 단어 13개, 용언 원형 2개)를 제외한 명사 단어 20,959개, 용언 원형 어절 1,080개 등 총 22,039개 어휘를, Kim과 Nam(2018), Kim 등(2020)에서는 중복되는 항목 164개(명사 어절 90개, 용언 어절 74개)를 제외한 명사 어절 480개, 용언 원형 어절 13개, 굴절형 어절 406개 등 총 899개 어휘에 대한 데이터를 추출하였다.

분석 방법

본 연구는 한국어 어휘의 시각 재인 시 나타나는 음절 빈도 효과를 재분석하고자 기존에 수집된 단어와 어절을 포함한 어휘 판단 과제 데이터를 기반으로 일관된 방식으로 첫음절 빈도를 산출하여 상관 및 회귀 분석을 진행하였다. 따라서 기존 연구에서 사용된 빈도 변인을 사용하지 않고 동일한 방식으로 빈도 변인을 재산출하였다. 먼저, 전체 형태 빈도는 각 어휘 항목의 세종 말뭉치상에 기재된 빈도를 추출하였다. 다음으로, 철자형 첫음절 빈도(orthographic first syllable frequency)는 어휘 항목의 첫음절이 가지는 철자 형태 원형을 기준으로, 음운형 음절 빈도(phonological first syllable frequency)는 표준 발음 규칙에 의거한 어휘 항목의 첫음절을 기준으로 계산되었다. 어휘 항목의 길이 제한 음절 빈도는 해당 어휘와 길이(음절 수)와 첫음절이 같은 이웃의 빈도로 계산되었으며, 길이를 제한하지 않은 음절 빈도 계산 방식은 어휘 항목의 길이와 무관하게 해당 첫음절로 시작하는 어휘를 이웃으로 간주하였다. 첫음절의 타입 빈도(type frequency)는 해당 음절로 시작하는 모든 어휘 항목의 수를 더한 값으로, 토큰 빈도(token frequency)는 해당 음절로 시작하는 모든 어휘 항목의 빈도를 누적한 값으로 계산하였다. 분석에 포함된 빈도 변인은 다음과 같다. 전체 형태 빈도, 길이 제한 철자형 첫음절 타입 빈도, 길이 무관 철자형 첫음절 타입 빈도, 길이 제한 철자형 첫음절 토큰 빈도, 길이 무관 철자형 첫음절 토큰 빈도, 길이 제한 음운형 첫음절 타입 빈도, 길이 무관 음운형 첫음절 타입 빈도, 길이 제한 음운형 첫음절 토큰 빈도, 길이 무관 음운형 첫음절 토큰 빈도.

이와 별도로, 음절 빈도 효과에 영향을 미치는 것으로 알려진 변인 또한 분석에 포함되었다. 먼저, 첫음절의 음운 변화 여부를 포함한 이유는 음운 변화 여부에 따른 음운형 빈도와 철자형 빈도의 효과의 차이를 조사하기 위함이었다. 예를 들어, 음운 변화가 없는 명사 단어 "전구"의 경우 철자형 첫음절과 음운형 첫음절이 "전"으로 동일하지만, 음운 변화가 나타나는 명사 단어 "전력"의 철자형 첫음절은 "전"인 반면, 음운형 첫음절은 "절"로 상이하게 나타난다. 이 경우 철자형 첫음절 "전"을 기준으로는 두 어휘가 이웃 관계이지만, 음운형 첫음절은 각각 "전"과 "절"로 상이하므로 더 이상 이웃 관계가 아니게 된다. 철자형, 음운형 첫음절 간의 일치 여부는 시각 어휘 재인에서 음절 빈도 효과에 유의한 영향을 미치며, 이러한 상호작용은 철자형 첫음절 빈도(Kwon et al., 2015), 음운형 첫음절 빈도(Tae et al., 2017) 모두에서 보고된 바 있다. 이에 본 연구는 철자형 음절과 음운형 음절 간의 일치 여부를 분석에 투입하여 단어, 어절을 포함한 어휘 항목에서 그 효과를 조사하고자 하였다.

다음으로, 첫음절의 의미 여부, 즉 첫음절이 독립적으로

4) 전체 형태 빈도를 포함하여, 후술할 모든 빈도값은 1500만 어절 규모 세종 말뭉치상의 빈도값을 15로 나눈 백만 어절당 출현빈도(occurrences per million)를 기준으로 하였으며, 표준화를 위하여 상용 로그 변환한 값을 사용하였다.

사용될 수 있는지 여부를 분석에 포함하였다. 첫음절이 의미를 가질 경우, 적어도 명사 단어에서 음절 빈도 효과가 나타나지 않았으며(Kwon & Lee, 2015), 저자들은 이 결과를 한자어 형태소에 의한 영향으로 해석하였다. 그러나 Jin 등(2018)에서는 명사 단어에 대한 어휘 판단 과제에서 단어 빈도, 두 철자 쌍 빈도, 형태소 빈도와 더불어 첫음절 한자 형태소의 의미 이웃 개수를 통제하였음에도 첫음절 빈도의 효과가 나타나지 않았으며, Kwon(2020)은 이에 대하여 첫음절 한자 형태소의 이웃 크기에 대한 통제뿐 아니라, 의미 요인을 배제한 음운 부호로만 이루어진 음절 사용이 필요함을 주장하기도 하였다. 이에 본 연구는 첫음절의 의미 여부에 따라 상이하게 나타나는 음절 빈도 효과가 모든 어휘 항목에서 동일한 패턴을 보이는지 여부를 조사하기 위하여 첫음절의 의미 여부를 분석에 포함하였다. 이때, 첫음절의 항목 중 최소한 한 개 이상이 의미를 가지는 경우 해당 첫음절은 의미를 가지는 것으로 간주하였다.

마지막으로, 어휘 항목의 음절 이웃 중 해당 어휘 항목보다 높은 빈도를 가진 항목의 수를 철자형 음절, 음운형 음절 각각에 대하여 계산하여 변수로 고려하였다. 이러한 고빈도 음절 이웃 수(higher frequency syllabic neighbors; 이하 HFSN)는 그 수가 많을 때 음절 빈도 효과가 억제적으로 나타나는 결과가 보고된 바 있다(Perea & Carreiras, 1998).

종합하면, 본 연구에서는 한국어 단어와 어절을 포함한 어휘의 시각 재인 과정에서 나타나는 음절 빈도 효과를 일관된 기준으로 분석하고, 그에 영향을 미치는 것으로 알려진 길이, 어휘 및 의미 요인을 분석에 포함하여 상이한 효과가 나타나는지를 조사하고자 하였다. 이를 위해 어휘 항목을 품사와 형태에 따라, 명사 단어, 명사 어절, 용언 원형, 용언 굴절형 네 가지 범주로 분류하고 상관 및 회귀 분석을 진행하였다. 4개 어휘 범주(명사 단어, 명사 어절, 용언 원형, 용언 굴절

Table 1. Descriptive statistics of characteristics and reaction time (ms) for simple and inflected words. *M* and *SD* are used to represent mean and standard deviation, respectively.

	Nouns (n = 21,439)				Verbs, adjectives, and adverbs (n = 1,499)			
	Simple (n = 20,959)		Inflected (n = 480)		Simple (n = 1,093)		Inflected (n = 406)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
LenSyll	2.47	0.7	3.28	0.62	3.52	0.84	3.3	0.69
Full-form Frequency	0	0.68	0.7	0.88	-0.81	0.51	0.58	0.88
HFOSN	23.01	25.34	6.56	10.99	12.07	18.55	5.19	9.43
HFPSN	24.51	27.33	7.09	11.87	15.53	22.21	6.83	11.44
Orthographic Type Frequency	2.48	0.54	2.59	0.48	2.31	0.6	2.4	0.57
Orthographic Token Frequency	3.26	0.63	3.43	0.53	3.16	0.68	3.35	0.61
Orthographic Type Frequency (Length-restraint)	1.48	0.62	1.93	0.48	1.68	0.66	1.75	0.59
Orthographic Token Frequency (Length-restraint)	2.64	0.73	2.92	0.53	2.55	0.66	2.77	0.63
Phonological Type Frequency	2.5	0.56	2.64	0.49	2.41	0.6	2.53	0.57
Phonological Token Frequency	3.3	0.67	3.5	0.58	3.28	0.7	3.46	0.63
Phonological Type Frequency (Length-restraint)	1.5	0.64	1.98	0.48	1.78	0.65	1.88	0.58
Phonological Token Frequency (Length-restraint)	2.69	0.77	2.98	0.57	2.66	0.68	2.88	0.66
RT	630	81.24	506.6	37.53	604.74	62.19	523.05	43.65

Note. All frequencies were transformed to per million values and statistical analyses were conducted on log-transformed values. *LenSyll* = word length in syllable; *HFOSN* = higher frequency orthographic syllabic neighbors; *HFPSN* = higher frequency phonological syllabic neighbors; *RT* = reaction time (ms).

형)별 길이 및 빈도 변인, 반응시간에 대한 기술 통계치가 Table 1에 제시되어 있다.

먼저 상관 분석에서는 상기 4개의 범주별로 평균 반응시간(reaction time, RT)과 범주 변인인 첫음절의 의미 여부, 음운 변화 여부를 제외한 7개 연속 변인, 즉 음절 수, 전체 형태 빈도, 고빈도 철자형 음절 이웃 수, 고빈도 음운형 음절 이웃 수, 철자형 타입 빈도, 철자형 토큰 빈도, 음운형 타입 빈도, 음운형 토큰 빈도, 길이 제한 철자형 타입 빈도, 길이 제한 철자형 토큰 빈도, 길이 제한 음운형 타입 빈도, 길이 제한 음운형 토큰 빈도 간의 피어슨 상관 계수(Pearson's R)를 조사하였다.

이어서 4개의 어휘 항목 범주 각각에 대하여 평균 반응시간을 종속 변인으로 하는 위계적 회귀 분석(hierarchical regression analysis)이 수행되었다. 선행된 상관 분석 결과 4개 어휘 범주 모두에서 본 연구의 주요 관심사인 8개 첫음절 빈도 계산 방식으로 산출된 변인들 간 상관이 높게 나타났으며, 분산팽창요인(variance inflation factors, VIF)로 측정된 다중공선성(multicollinearity)이 기준치를 초과하였기 때문에, 회귀 모형별로 하나의 첫음절 빈도 계산 방식만을 투입하여 각각의 효과 크기를 비교하고자 하였다. 또한, 고빈도 철자형 음절 이웃 수와 고빈도 음운형 음절 이웃 수 또한 4개 어휘 범주 모두에서 상관 및 다중공선성이 상당히 높게 나타나(명사 단어: r = .94, p 〈 .001; 명사 어절: r = .97, p 〈 .001; 용언 원형: r = .85, p 〈 .001; 용언 굴절형: r = .84, p 〈 .001), 회귀 분석에서는 첫음절 빈도 계산 방식 중 철자형 음절을 기준으로 한 변인의 경우 고빈도 철자형 음절 이웃 수를, 음운형 음절을 기준으로 한 변인의 경우 고빈도 음운형 음절 이웃 수를 분석에 투입하였다(자세한 변인 간 상관 및 VIF 수치는 부록 I, 부록 II 참고).

따라서 어휘 판단 시간에 대한 위계적 회귀 분석의 1단계 모형에서는 6개의 예측 변인 중 전체 형태 빈도와 어휘 항목의 길이(음절 수)를, 2단계 모형에서는 첫음절의 의미, 음운 변화 여부 및 고빈도 이웃 수를 투입하였으며, 마지막 3단계 모형에서는 8개 첫음절 빈도 계산 방식 중 하나의 첫음절 빈도 변인만을 투입하여 분석을 진행하였다. 단계별로 각 요인 군집(factor cluster)이 투입된 이유는 다음과 같다. 먼저, 1단계에서 다수의 시각 단어 재인 연구에서 공통적으로 보고되거나 고려되는 변인들을 포함한 어휘 요인 군집(cluster)의 영향을, 2단계에서는 첫음절의 속성을 나타내는 하위 어휘 요인 군집의 영향을 통제하고자 하였다. 이후 최종 3단계에서는 표면 변인(surface variable) 및 어휘 변인(lexical variable)을 통제한 후의 의미 변인의 영향력을 조사한 Yap과 Balota(2009)의 위계적 회귀 연구와 마찬가지로, 앞선 단계의 요인 군집을 통제한 후 음절 빈도 효과를 조사하고자 하였다. 전체 형태 빈도, 8개 첫음절 빈도를 포함한 빈도 변인은 모두 세종 말뭉치 상에서 백만 어절당 출현 횟수(occurrences per million)를 로그 변환(log-transformation)한 값을 사용하였으며, 어휘 판단 시간 역시 표준화를 위하여 다른 빈도 변인과 마찬가지로 로그 변환한 값을 사용하였다.

결 과

한국어 시각 어휘 재인에서 첫음절 빈도의 효과를 조사하기 위하여, 네 가지 어휘 범주 각각에 대하여 여러 종류의 첫음절 빈도 변인과 어휘 판단 시간 간의 단순 상관 분석 결과는 부록 I에 제시되어 있으며, 철자형 및 음운형 음절 빈도 각 4개의 독립적인 위계적 회귀 모형 분석을 실시한 결과는 Table 2, Table 4에 제시되어 있다. 단순 상관 분석은 해당 변인과 관련된 다른 변인들과의 설명 변량도 포함하고 있어, 해당 변인의 독자적인 설명량을 추정할 수 없고, 따라서 위계적 회귀 분석을 통하여 해당 변인의 독자적 설명량의 유의미성을 평가하고자 하였다. 위계적 회귀 모형에는 단순 상관 분석에서 어휘 판단 시간과 유의한 상관관계를 보인 첫음절 빈도 변인을 대상으로 하였다. 또한, 위계적 회귀 분석은 사례 수에 의해 영향을 받을 수가 있기 때문에 이러한 제한점을 보완하고자 단순 상관 분석 역시 상호보완적으로 활용하여 결과 논의를 진행하였다. 이후의 결과 분석에서 단순 상관과 위계적 회귀 분석 모두에서 유의한 결과를 보인 첫음절 빈도 변인 위주로 결과를 논의하였다. 위계적 회귀 분석은 1, 2단계에서 첫음절 빈도와 함께 종속 변인과 관련이 있는 것으로 알려진 변인들의 효과를 통제한 뒤 3단계에서 최종적으로 투입한 8개 첫음절 빈도 변인 각각의 설명력(수정된 R 제곱, adjusted R squared, 이하 R^2_{adj}) 변화량 비교를 통해 각 첫음절 빈도 계산 방식의 어휘 판단 시간에 대한 설명력과 효과 크기를 비교 분석하고자 하였다.

철자형 음절 분석 결과

네 어휘 범주에 대하여 타입, 토큰 및 길이 제한 여부에 따른 4개의 음절 빈도 계산 방식을 마지막 단계의 예측 변인으로 하는 각 4개의 독립적인 위계적 회귀 분석을 수행한 결과는 Table 4에 제시되어 있으며, 최종 3단계 모형은 Table 5에 제시되어 있다.

명사 어휘에서는 단어, 어절에 따라 철자형 첫음절 빈도의

Table 2. Regression coefficients from Step 1 to 3 of the item-level hierarchical regression analyses for lexical decision times (log-transformed RT, milliseconds) for simple and inflected words. In Step 3, one of the four frequency measures for orthographic syllable frequency of first syllable was entered in each model (1-1 – 1-4).

Predictor variable	Nouns (n = 21,439)		Verbs, adjectives, and adverbs (n = 1,499)	
	Simple (n = 20,959)	Inflected (n = 480)	Simple (n = 1,093)	Inflected (n = 406)
Step 1: Lexical Variables				
Word length in syllable	−.017***	−.002	−.010***	−.004
Full-form frequency	−.031***	−.017***	−.016***	−.011***
R^2_{adj}	**.162*** **	**.199*** **	**.046*** **	**.063*** **
Step 2: First Syllable characteristics				
Meaning	.002	−.004	−.007	.008*
Phonological change	−.002†	.012**	−.003	.011**
HFOSN	.000***	−.000*	−.000	−.000*
R^2_{adj}	**.163*** **	**.221*** **	**.045*** **	**.100*** **
	ΔR^2 = .022* **	**ΔR^2 = .001*** **	**ΔR^2 = −.001**	**ΔR^2 = .037*** **
Step 3: Orthographic Syllable Frequency				
1-1. Orthographic Type Frequency	.001	−.013***	−.017***	−.022***
R^2_{adj}	**.163*** **	**.241*** **	**.065*** **	**.159*** **
	ΔR^2 = .000	**ΔR^2 = .020*** **	**ΔR^2 = .020*** **	**ΔR^2 = .059*** **
1-2. Orthographic Token Frequency	−.000	−.012***	−.015***	−.018***
R^2_{adj}	**.163*** **	**.243*** **	**.071*** **	**.152*** **
	ΔR^2 = .000	**ΔR^2 = .021*** **	**ΔR^2 = .026*** **	**ΔR^2 = .052*** **
1-3. Orthographic Type Frequency (fixed-length)	−.005***	−.014***	−.021***	−.025***
R^2_{adj}	**.164*** **	**.243*** **	**.079*** **	**.176*** **
	ΔR^2 = .001* **	**ΔR^2 = .021*** **	**ΔR^2 = .034*** **	**ΔR^2 = .076*** **
1-4. Orthographic Token Frequency (fixed-length)	−.002**	−.014***	−.021***	−.018***
R^2_{adj}	**.163*** **	**.252*** **	**.097*** **	**.155*** **
	ΔR^2 = .000 **	**ΔR^2 = .031*** **	**ΔR^2 = .052*** **	**ΔR^2 = .055*** **

Note. All frequencies were transformed to per million values and statistical analyses were conducted on log-transformed values.
R^2_{adj} = adjusted R square; *HFOSN* = higher frequency orthographic syllable neighbors.
† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

상이한 효과가 나타났다. 명사 단어에서는 철자형 음절 빈도 4가지 계산 방식의 3단계 모형은 모두 유의하였으나 설명력이 높지 않았으며(타입 빈도: R^2_{adj} = .000, F(6, 20952)= 682.7; 토큰 빈도: R^2_{adj} = .000, F(6, 20952)= 682.6; 길이 제한 타입 빈도: R^2_{adj} = .001, F(6, 20952)= 686.6; 길이 제한 토큰 빈도: R^2_{adj} = .000, F(6, 20952)= 684.4; all ps < .001; Table 3 참고), 상관 분석 결과(부록 1-1)에서도 반응시간과 예측 변인 간의 유의하지만 낮은 상관 계수(타입 빈도: r = −.04; 토큰 빈도: r = −.09; all ps < .001)가 보고되었다. Table 2에 나타난 바와 같이, 길이 무관 음절 빈도는 유의하지 않게 나타났으며(타입 빈도: ΔR^2_{adj} = .000; β = .001, p = .47; 토큰 빈도: ΔR^2_{adj} = .000, β = −.000, p = .68), 길이 제한 음절 빈도는 유의한 결과를 보였으나 어휘 판단 시간에 대한 설명력이 높지 않게 나타났다

Table 3. Regression coefficients of final models in hierarchical regression analyses for orthographic first syllable frequency

Predictor variable	Nouns		Verbs, adjectives, and adverbs	
	Simple	Inflected	Simple	Inflected
Model 1-1. Orthographic Type Frequency				
Word length in syllable	−.017***	.002	−.004*	.002
Full-form frequency	−.03***	−.015***	−.011***	−.005*
First syllable meaning	.003	−.01†	−.018**	.004
Phonological change	−.002†	.007†	−.009*	.001
HFOSN	0**	0	0*	0
Orthographic Type Frequency	.001	−.013***	−.017***	−.022***
Fit	R^2_{adj} = .163***	R^2_{adj} = .241***	R^2_{adj} = .065***	R^2_{adj} = .159***
Model 1-2. Orthographic Token Frequency				
Word length in syllable	−.016***	.002	−.004*	.002
Full-form frequency	−.03***	−.015***	−.009***	−.006*
First syllable meaning	.002	−.007	−.017**	.006†
Phonological change	−.002*	.008*	−.007*	.006
HFOSN	0**	0	0*	0
Orthographic Token Frequency	.000	−.012***	−.015***	−.018***
Fit	R^2_{adj} = .163***	R^2_{adj} = .243***	R^2_{adj} = .071***	R^2_{adj} = .152***
Model 1-3. Orthographic Type Frequency (fixed-length)				
Word length in syllable	−.014***	.004	.001	.008*
Full-form frequency	−.028***	−.015***	−.011***	−.005*
First syllable meaning	−.001	−.01†	−.02***	.003
Phonological change	−.003**	.006	−.009**	0
HFOSN	0**	0	0**	0
Orthographic Type Frequency	−.005***	−.014***	−.021***	−.025***
Fit	R^2_{adj} = .164***	R^2_{adj} = .243***	R^2_{adj} = .079***	R^2_{adj} = .176***
Model 1-4. Orthographic Token Frequency (fixed-length)				
Word length in syllable	−.016***	−.001	−.006***	−.002
Full-form frequency	−.029***	−.014***	−.007*	−.005†
First syllable meaning	0	−.008	−.018***	.006†
Phonological change	−.003*	.007†	−.008*	.006
HFOSN	0**	0	0**	0
Orthographic Token Frequency	−.002**	−.014***	−.021***	−.018***
Fit	R^2_{adj} = .163***	R^2_{adj} = .252***	R^2_{adj} = .097***	R^2_{adj} = .155***

Note. All frequencies were transformed to per million values and statistical analyses were conducted on log-transformed values. R^2_{adj} = adjusted R square; *HFOSN* = higher frequency orthographic syllable neighbors.
† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

(길이 제한 타입 빈도: ΔR^2_{adj} = .001, β = −.005; 길이 제한 토큰 빈도: ΔR^2_{adj} = .000, β = −.002; all ps 〈 .001).

반면, 명사 어절에서는 네 가지 철자형 음절 빈도 모두 반응시간의 감소를 유의미하게 설명했으며, 특히 길이 제한 토큰 빈도의 영향이 가장 크게 나타났다(타입 빈도: ΔR^2_{adj} = .020; β = .013; 토큰 빈도: ΔR^2_{adj} = .021, β = −.012; 길이 제한 타입 빈도: ΔR^2_{adj} = .021, β = −.014; 길이 제한 토큰 빈도: ΔR^2_{adj} = .031, β = −.014; all ps 〈 .001; Table 2 참고). 또한, 네 모형 모두 유의하였으나(타입 빈도: R^2_{adj} = .020, F(6, 473)= 26.3; 토큰 빈도: R^2_{adj} = .021, F(6, 473)= 26.62; 길이 제한 타입 빈도: R^2_{adj} = .021, F(6, 473)= 26.56; 길이 제한 토큰 빈도: R^2_{adj} = .031, F(6, 473)= 27.83; all ps 〈 .001; Table 3 참고), 그 중 길이 제한 토큰 빈도를 사용한 모형의 설명력이 가장 높게 나타났다(R^2_{adj} = .031).

용언에서는 원형, 굴절형 모두 철자형 음절 빈도와 어휘 판단 시간 간의 유의한 부적 관계가 나타났다. 굴절형의 네 가지 모형(타입 빈도: R^2_{adj} = .159, F(6, 399)= 13.79; 토큰 빈도: R^2_{adj} = .152, F(6, 399)= 13.13; 길이 제한 타입 빈도: R^2_{adj} = .176, F(6, 399)= 15.38; 길이 제한 토큰 빈도: R^2_{adj} = .155, F(6, 399)= 13.36; all ps 〈 .001; Table 3 참고)에서 전반적으로 설명력이 원형(타입 빈도: R^2_{adj} = .065, F(6, 1086)= 13.67; 토큰 빈도: R^2_{adj} = .071, F(6, 1086)= 14.98; 길이 제한 타입 빈도: R^2_{adj} = .079, F(6, 1086)= 16.54; 길이 제한 토큰 빈도: R^2_{adj} = .097, F(6, 1086)= 20.46; all ps 〈 .001; Table 3 참고) 대비 상대적으로 높게 나타났다. Table 2에 나타난 바와 같이, 원형에서는 길이 제한 토큰 빈도(ΔR^2_{adj} = .052, β = −.021, p 〈 .001)가 가장 큰 효과를 보였지만, 굴절형에서는 길이 제한 타입 빈도(ΔR^2_{adj} = .076, β = −.018, p 〈 .001)가 설명력 변화가 크게 나타났다.

철자형 첫음절 빈도를 사용한 위계적 회귀 분석 결과, 네 가지 어휘 범주 중 명사 단어를 제외한 명사 어절, 용언 원형 및 굴절형 모두에서 철자형 첫음절 빈도의 촉진적 효과가 발견되었으며, 이는 어휘 변인과 더불어 음절 빈도 효과에 영향을 미치는 첫음절의 속성을 통제한 뒤에도 유의한 것으로 나타났다.

음운형 음절 분석 결과

4개 어휘 범주의 어휘 판단 시간에 대한 네 가지 음운형 음절 빈도 계산 방식에 따른 위계적 회귀 모형 분석 결과는 Table 4에 제시되어 있으며, 최종 모형은 Table 5에 제시되어 있다. 철자형 음절 빈도와 마찬가지로, 명사 단어를 제외한 명사 어절, 용언 원형 및 굴절형 모두에서 첫음절 빈도의 촉진적 효과가 나타났다. 명사 단어에서는 네 가지 음운형 음절 빈도 계산 방식에 따른 모형 모두 유의하였으나(타입 빈도: R^2_{adj} = .163, F(6, 20952)= 681.4; 토큰 빈도: R^2_{adj} = .163, F(6, 20952)= 681.4; 길이 제한 타입 빈도: R^2_{adj} = .164, F(6, 20952)= 685.9; 길이 제한 토큰 빈도: R^2_{adj} = .163, F(6, 20952)= 683.4; all ps 〈 .001; Table 5 참고), 길이 제한 타입 빈도와 토큰 빈도에서 반응 시간과 약간의 부적 관계를 보였을 뿐(길이 제한 타입 빈도: ΔR^2_{adj} = .001, β = −.005, p 〈 .001; 길이 제한 토큰 빈도: ΔR^2_{adj} = .000, β = −.002, p 〈 .001; Table 4 참고), 그 영향이 크지 않았다. 철자 빈도에서와 마찬가지로 상관 분석 결과(부록 I-1) 이와 유사하게 유의하지만 높지 않은 상관 계수(타입 빈도: r = −.04; 토큰 빈도: r = −.09; 길이 제한 타입 빈도: r = −.05; 길이 제한 토큰 빈도: r = −.07; all ps 〈 .001).

반면 명사 어절에서는 네 개의 음절 빈도 변인 모두 반응시간의 감소를 유의미하게 설명하였으며(타입 빈도: ΔR^2_{adj} = .029, β = −.014; 토큰 빈도: ΔR^2_{adj} = .026, β = −.012; 길이 제한 타입 빈도: ΔR^2_{adj} = .032, β = −.014; 길이 제한 토큰 빈도: ΔR^2_{adj} = .036, β = −.014, all ps 〈 .001; Table 4 참고), 4개 모형 모두 유의하였다(타입 빈도: R^2_{adj} = .25, F(6, 473)= 27.23; 토큰 빈도: R^2_{adj} = .247, F(6, 473)= 27.23; 길이 제한 타입 빈도: R^2_{adj} = .253, F(6, 473)= 28.03; 길이 제한 토큰 빈도: R^2_{adj} = .257, F(6, 473)= 28.66; all ps 〈 .001; Table 5 참고).

용언 원형 및 굴절형 역시 모든 음운형 첫음절 빈도에서 반응시간에 대한 유의한 촉진 효과가 발견되었으며, 모든 모형이 유의한 결과를 보고하였다(Table 5 참고). 철자형 음절 빈도의 회귀 분석 결과와 마찬가지로, 용언 원형의 네 모형(타입 빈도: R^2_{adj} = .051, F(6, 1086)= 10.79; 토큰 빈도: R^2_{adj} = .055, F(6, 1086)= 11.63; 길이 제한 타입 빈도: R^2_{adj} = .064, F(6, 1086)= 13.5; 길이 제한 토큰 빈도: R^2_{adj} = .070, F(6, 1086)= 14.71; all ps 〈 .001)은 굴절형의 네 모형(타입 빈도: R^2_{adj} = .165, F(6, 399)= 14.35; 토큰 빈도: R^2_{adj} = .167, F(6, 399)= 14.56; 길이 제한 타입 빈도: R^2_{adj} = .171, F(6, 399)= 14.97; 길이 제한 토큰 빈도: R^2_{adj} = .166, F(6, 399)= 14.4; all ps 〈 .001) 대비 상대적으로 설명력이 낮게 나타났다. Table 4에 나타난 바와 같이, 원형에서는 길이 제한 토큰 빈도의 영향이 가장 크게 나타났으며(ΔR^2_{adj} = .035, β = −.021, p 〈 .001), 굴절형

Table 4. Regression coefficients from Step 1 to 3 of the item-level hierarchical regression analyses for lexical decision times (log-transformed RT, milliseconds) for simple and inflected words. In Step 3, one of the four frequency measures (log-transformed, per million) for phonological syllable frequency of fist syllable was entered in each model (5 - 8).

Predictor variable	Nouns (n = 21,439)		Verbs, adjectives, and adverbs (n = 1,499)	
	Simple (n = 20,959)	Inflected (n = 480)	Simple (n = 1,093)	Inflected (n = 406)
Step 1: Lexical Variables				
Word length in syllable	−.017***	−.002	−.010***	−.004
Full-form frequency	−.031***	−.017***	−.016***	−.011***
R^2_{adj}	**.162*** **	**.199*** **	**.046*** **	**.063*** **
Step 2: Syllable 1 characteristics				
Meaning	.002	−.004	−.007	.008*
Phonological change	−.003**	.012**	−.003	.015***
HFPSN	.000***	−.000*	−.000	−.000***
R^2_{adj}	**.163*** **	**.221*** **	**.045*** **	**.125*** **
	ΔR^2 = .001* **	**ΔR^2 = .022** **	**ΔR² = −.001**	**ΔR^2 = .062*** **
Step 3: Syllable 1 frequency				
2-1. Phonological Type Frequency	.000	−.014***	−.009**	−.018***
R^2_{adj}	**.163*** **	**.25*** **	**.051*** **	**.165*** **
	ΔR^2 = .000	**ΔR^2 = .029*** **	**ΔR^2 = .006** **	**ΔR^2 = .040*** **
2-2. Phonological Token Frequency	−.000	−.012***	−.009***	−.016***
R^2_{adj}	**.163*** **	**.247*** **	**.055*** **	**.167*** **
	ΔR^2 = .000	**ΔR^2 = .026*** **	**ΔR^2 = .010*** **	**ΔR^2 = .042*** **
2-3. Phonological Type Frequency (fixed-length)	−.005***	−.014***	−.015***	−.019***
R^2_{adj}	**.164*** **	**.253*** **	**.064*** **	**.171*** **
	ΔR^2 = .001* **	**ΔR^2 = .032*** **	**ΔR^2 = .019*** **	**ΔR^2 = .046*** **
2-4. Phonological Token Frequency (fixed-length)	−.002**	−.014***	−.021***	−.015***
R^2_{adj}	**.163*** **	**.257*** **	**.070*** **	**.166*** **
	ΔR^2 = .000 **	**ΔR^2 = .036*** **	**ΔR^2 = .035*** **	**ΔR^2 = .041*** **

Note. All frequencies were transformed to per million values and statistical analyses were conducted on log-transformed values. R^2_{adj} = adjusted R square; *HFPSN* = higher frequency phonological syllable neighbors.
† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

에서는 길이 제한 타입 빈도의 영향이 가장 크게 나타났다 (ΔR^2_{adj} = .046, β = −.025, $p < .001$).

음운형 음절 빈도에 대한 위계적 회귀 분석 결과는 철자형 음절 빈도의 분석 결과와 유사한 패턴을 보였으며, 최종 3단계 모형에서 명사 단어를 제외한 세 어휘 범주에서 첫음절 음운형 빈도의 유의한 촉진 효과가 나타났다.

종합하면, 철자형, 음운형 첫음절 빈도 모두에서 명사 단어를 제외한 명사 어절, 용언 원형 및 굴절형에서 반응시간에 대한 촉진적 효과를 보였다. 또한, 명사 어절에서는 철자형 음절 빈도 대비 음운형 음절 빈도의 설명력이 높았던 반면(e.g., 음운형 길이 제한 토큰 빈도 모형: R^2_{adj} = .257, ΔR^2_{adj} = .036, β = −.014, $p < .001$), 용언 원형 및 굴절형에서는 철자형 음절 빈도의 설명력이 상대적으로 높게 나타났다(e.g., 원형의 철자형 길이 제한 토큰 빈도 모형: R^2_{adj} = .097, ΔR^2_{adj} = .052, β = −.021, $p < .001$; 굴절형의 철자형 길이 제한 타입 빈도 모형: R^2_{adj} = .176, ΔR^2_{adj} =

Table 5. Regression coefficients of final models in hierarchical regression analyses for phonological first syllable frequency

Predictor variable	Nouns		Verbs, adjectives, and adverbs	
	Simple	Inflected	Simple	Inflected
Model 1-1. Phonological Type Frequency				
Word length in syllable	−.017***	.002	−.006**	0
Full-form frequency	−.03***	−.013***	−.014***	−.008**
First syllable meaning	.002	−.009†	−.011*	.005
Phonological change	−.003**	.011**	−.003	.013**
HFPSN	0*	0	0	0
Phonological Type Frequency	0	−.015***	−.009**	−.018***
Fit	R^2_{adj} = .163***	R^2_{adj} = .25***	R^2_{adj} = .051***	R^2_{adj} = .165***
Model 1-2. Phonological Token Frequency				
Word length in syllable	−.017***	.002	−.006**	0
Full-form frequency	−.03***	−.014***	−.012***	−.008**
First syllable meaning	.002	−.006	−.011*	.007*
Phonological change	−.003**	.013***	−.003	.014***
HFPSN	0**	0	0	0
Phonological Token Frequency	0	−.012***	−.009***	−.016***
Fit	R^2_{adj} = .163***	R^2_{adj} = .247***	R^2_{adj} = .055***	R^2_{adj} = .167***
Model 1-3. Phonological Type Frequency (fixed-length)				
Word length in syllable	−.013***	.005	−.002	.005
Full-form frequency	−.028***	−.014***	−.012***	−.009**
First syllable meaning	−.001	−.01†	−.014*	.005
Phonological change	−.004**	.01**	−.003	.014***
HFPSN	0**	0	0*	0
Phonological Type Frequency	−.005***	−.016***	−.015***	−.019***
Fit	R^2_{adj} = .164***	R^2_{adj} = .253***	R^2_{adj} = .064***	R^2_{adj} = .171***
Model 1-4. Phonological Token Frequency (fixed-length)				
Word length in syllable	−.016***	−.001	−.006***	−.003
Full-form frequency	−.029***	−.012***	−.01***	−.008**
First syllable meaning	0	−.007	−.012*	.007*
Phonological change	−.003**	.011**	−.003	.015***
HFPSN	0**	0	0**	0
Phonological Token Frequency	−.002**	−.014***	−.014***	−.015***
Fit	R^2_{adj} = .163***	R^2_{adj} = .257***	R^2_{adj} = .070***	R^2_{adj} = .166***

Note. All frequencies were transformed to per million values and statistical analyses were conducted on log-transformed values.
R^2_{adj} = adjusted R square; *HFPSN* = higher frequency phonological syllable neighbors.
† $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

.076, β = −.025, p < .001).

선형 혼합 효과 모형 분석 결과

위계적 회귀 분석과 더불어, 본 연구는 선형 혼합 효과 모형(linear mixed-effect model)을 사용하여 네 어휘 범주에서 각 변인이 음절 빈도 효과에 미치는 영향을 조사하고자 변인 간의 상호작용을 포함한 추가 분석을 진행하였다. 선형 혼합 효과 모형 분석을 위하여 R 프로그램(version 4.2.1, R core team, 2022)의 nlme 패키지(Pinheiro et al., 2017)를 사용하였고, 분석 결과의 시각화를 위하여 ggplot2 패키지(Wickham, 2009)를 사용하였다. 위계적 다중 회귀 분석 모형과 마찬가지로 네 어휘 범주 모두에서 철자형 및 음운형 음절 빈도, 길이 제한 및 길이 무관 빈도에 대하여 8개의 독립적인 모형을 구축하였다. 또한, 단어 길이, 첫음절의 의미 유무, 음운 변동 여부는 R 프로그램의 factor function을 사용하여 요인화한 후 고정 변인으로 투입하였으며, 단어 빈도, 고빈도 음절 이웃 수, 음절 빈도는 요인화하지 않은 연속 변인으로서 고정 변인으로 투입되었다. 청각 제시된 단어의 길이(duration)을 제외한 단어 빈도, 형태소 수, 구체성 등 모든 변인과 과제 간의 상호작용항을 선형 혼합 효과 모형의 고정 변인으로 투입한 Goh 등(2016)의 분석 절차와 마찬가지로, 단어 길이를 제외한 모든 어휘 변인 및 첫음절 변인과 음절 빈도 간의 상호작용항이 고정 변인으로 모형에 포함되었다. Yi 등(2017)의 어휘 판단 데이터에는 모든 참가자의 반복 측정 데이터가 아닌 어휘 항목별 평균 반응시간 데이터만 접근 가능하였기 때문에, 어휘 항목을 무선 변인으로 투입하였다. 철자 음절 빈도에 대한 모형은 (1)과 같이 설정되었으며, 음운 음절 빈도에 대한 모형은 (2)와 같이 설정되었다.

RT ~ WL + (first_meaning + phon_change + HFSN_O + WF) * SF_O + (1|item) (1)

RT ~ WL + (first_meaning + phon_change + HFSN_P + WF) * SF_P + (1|item) (2)

위 표현에서 WL은 음절 수를 기준으로 한 단어 길이, first_meaning은 첫음절의 의미 여부, phon_change는 첫음절의 음운 변동 여부, SF_O는 철자 음절 빈도, SF_P는 음운 음절 빈도, HFSN_O는 철자 음절 이웃 수, HFSN_P는 음운 음절 이웃 수, WF는 단어 빈도, item은 어휘 항목을 의미한다.

명사 단어에 대한 선형 혼합 효과 모형 분석 결과(Table 6 좌측 참고), 철자형 음절에 대한 분석에서는 위계적 회귀 모형 분석 결과와 마찬가지로 길이 무관 타입 및 길이 무관 토큰 빈도는 모두 유의한 주효과를 보이지 않았으나(길이 무관 타입 빈도: β = 0, SE = .001, t = 0, p = .999; 길이 무관 토큰 빈도: β = 0, SE = .001, t = −0.21, p = .836), 길이 제한 음절 빈도는 타입 및 토큰 빈도 모두에서 유의한 촉진 효과가 발견되었다(길이 제한 타입 빈도: β = −.004, SE = .001, t = −3.46, p < .001; 길이 제한 토큰 빈도: β = −.002, SE = .001, t = −3.13, p = .002). 또한, 음절 수를 기준으로 한 단어 길이와 단어 빈도는 모두 촉진적 효과를 보였지만, 고빈도 음절 이웃 수는 모든 철자형 음절 빈도 모형에서, 첫음절의 의미 유무는 길이 제한 타입 빈도를 제외한 모든 빈도 모형에서 억제적 효과를 보였다. 첫음절의 음운 변동 여부는 길이 제한 토큰 빈도 모형에서만 유의한 촉진 효과를 나타냈다.

모든 철자형 음절 빈도 모형에서 음절 빈도와 고빈도 음절 이웃 수 간의 상호작용(길이 무관 타입 빈도: β = 0, SE = 0, t = −3.14, p = .002; 길이 무관 토큰 빈도: β = 0, SE = 0, t = −3.79, p < .001; 길이 제한 타입 빈도: β = 0, SE = 0, t = −11.02, p < .001; 길이 제한 토큰 빈도: β = 0, SE = 0, t = −5.53, p < .001)과 음절 빈도와 첫음절의 의미 유무 간의 상호작용이 유의하게 나타났다(길이 무관 타입 빈도: β = −.009, SE = .003, t = −2.88, p = .004; 길이 무관 토큰 빈도: β = −.008, SE = .003, t = −2.99, p = .003; 길이 제한 타입 빈도: β = −.008, SE = .003, t = −3.01, p = .003; 길이 제한 토큰 빈도: β = −.007, SE = .002, t = −3.06, p = .002). 단어 빈도와 음절 빈도 간의 상호작용은 길이 제한 타입 빈도를 제외한 모든 모형에서 유의하게 나타났으며(길이 무관 타입 빈도: β = .003, SE = .001, t = 3.23, p = .001; 길이 무관 토큰 빈도: β = .004, SE = .001, t = 4.8, p < .001; 길이 제한 토큰 빈도: β = .004, SE = .001, t = 4.81, p < .001), 첫음절의 음운 변동 여부와 음절 빈도 간의 상호작용은 모든 모형에서 유의하지 않았다.

명사 단어에 대한 음운형 음절 빈도의 선형 혼합 효과 모형 분석 또한 철자형 음절 빈도와 유사한 결과 패턴을 보였으며(Table 6 우측 참고), 길이 제한 타입 및 토큰 빈도 모형에서 음절 빈도의 주효과가 촉진적으로 나타났다(길이 제한 타입 빈도: β = −.004, SE = .001, t = −3.97, p < .001; 길이 제한 토큰 빈도: β = −.002, SE = .001, t = −2.71, p = .007). 다음으로, 모든 음운형 음절 빈도 모형에서 단어 길이와 단어 빈도는 촉진적 효과를 보였지만, 첫음

Table 6. Results of linear mixed-effect model of reaction time data for simple nouns

Fixed effects	Orthographic syllable			Phonological syllable		
	β	*SE*	*t*-value	β	*SE*	*t*-value
Model 1. Type Frequency						
(Intercept)	2.83	.002	1139.94***	2.83	.002	1213.55***
Word length in syllable	−.016	.001	−23.14***	−.016	.001	−23.08***
Word frequency	−.037	.003	−14.11***	−.034	.003	−13.57***
First syllable meaning	.017	.005	3.06**	.015	.005	2.65**
Phonological change	−.004	.004	−0.97	−.001	.004	−0.32
HFSN	.001	0	3.59***	.001	0	3.9***
Syllable Frequency	0	.001	0	0	.001	−0.08
Word frequency * Syllable Frequency	.003	.001	3.23**	.002	.001	2.29*
First syllable meaning * Syllable Frequency	−.009	.003	−2.88**	−.007	.003	−2.44*
Phonological change * Syllable Frequency	.001	.002	0.51	−.001	.002	−0.34
HFSN * Syllable Frequency	0	0	−3.14**	0	0	−3.54***
Model 2. Token Frequency						
(Intercept)	2.83	.003	1055.1***	2.83	.002	1137.87***
Word length in syllable	−.015	.001	−25.3***	−.016	.001	−25.9***
Word frequency	−.042	.003	−14.31***	−.038	.003	−13.71***
First syllable meaning	.02	.006	3.12**	.019	.006	3.01**
Phonological change	−.004	.005	−0.79	−.006	.005	−1.23
HFSN	.001	0	4.52***	.001	0	4.61***
Syllable Frequency	0	.001	−0.21	0	.001	0.06
Word frequency * Syllable Frequency	.004	.001	4.8***	.003	.001	3.48***
First syllable meaning * Syllable Frequency	−.008	.003	−2.99**	−.007	.002	−2.84**
Phonological change * Syllable Frequency	.001	.002	0.37	.001	.001	0.72
HFSN * Syllable Frequency	0	0	−3.79***	0	0	−4.02***
Model 3. Type Frequency (fixed-length)						
(Intercept)	2.82	.002	1480.09***	2.82	.002	1485.7***
Word length in syllable	−.01	.001	−10.45***	−.009	.001	−10.23***
Word frequency	−.024	.001	−18.64***	−.024	.001	−18.8***
First syllable meaning	.005	.003	1.79†	.005	.003	1.61
Phonological change	−.002	.002	−1	−.002	.002	−1
HFSN	.001	0	12.74***	.001	0	13.13***
Syllable Frequency	−.004	.001	−3.46***	−.004	.001	−3.97***
Word frequency * Syllable Frequency	−.001	.001	−1.59	−.002	.001	−1.86†
First syllable meaning * Syllable Frequency	−.008	.003	−3.01**	−.007	.003	−2.57*
Phonological change * Syllable Frequency	−.001	.002	−0.38	0	.002	−0.04
HFSN * Syllable Frequency	0	0	−11.02***	0	0	−11.56***
Model 4. Token Frequency (fixed-length)						
(Intercept)	2.83	.002	1411.81***	2.83	.002	1485.77***
Word length in syllable	−.014	.001	−22.66***	−.014	.001	−23.26***
Word frequency	−.037	.002	−17.76***	−.035	.002	−17.41***
First syllable meaning	.012	.004	2.76**	.011	.004	2.58**
Phonological change	−.008	.004	−2.25*	−.008	.003	−2.25*
HFSN	.001	0	6.93***	.001	0	7.4***
Syllable Frequency	−.002	.001	−2.51*	−.002	.001	−2.71**
Word frequency * Syllable Frequency	.004	.001	4.81***	.003	.001	3.76***
First syllable meaning * Syllable Frequency	−.007	.002	−3.06**	−.006	.002	−2.74**
Phonological change * Syllable Frequency	.003	.002	1.66†	.002	.001	1.7†
HFSN * Syllable Frequency	0	0	−5.53***	0	0	−6.12***

Note. All frequencies were transformed to per million values and statistical analyses were conducted on log-transformed values. *HFSN* = higher frequency syllabic neighbors.
† $p < .1$ * $p < .05$ ** $p < .01$ *** $p < .001$

절의 의미 유무와 고빈도 음절 이웃 수는 억제적 효과를 보였다.

또한, 음절 빈도와 고빈도 음절 이웃 수 간의 상호작용(길이 무관 타입 빈도: β = 0, SE = 0, t = −3.54, p < .001; 길이 무관 토큰 빈도: β = 0, SE = 0, t = −4.02, p < .001; 길이 제한 타입 빈도: β = 0, SE = 0, t = −11.56, p < .001; 길이 제한 토큰 빈도: β = 0, SE = 0, t = −6.12, p < .001)과 더불어, 음절 빈도와 첫음절의 의미 유무 간의 상호작용(길이 무관 타입 빈도: β = .003, SE = .001, t = 2.45, p = .01; 길이 무관 토큰 빈도: β = .004, SE = .001, t = 3.99, p < .001; 길이 제한 타입 빈도: β = −.002, SE = .001, t = −2.53, p = .01; 길이 제한 토큰 빈도: β = .003, SE = .001, t = 3.51, p < .001)이 모두 유의하게 나타났다. 단어 빈도와 음절 빈도 간의 상호작용은 길이 제한 타입 빈도를 제외한 모든 모형에서 유의하게 나타났다(길이 무관 타입 빈도: β = 003, SE = .001, t = 3.23, p = .001; 길이 무관 토큰 빈도: β = .004, SE = .001, t = 4.8, p < .001; 길이 제한 토큰 빈도: β = .004, SE = .001, t = 4.81, p < .001).

명사 단어에 대한 어휘 판단 데이터의 선형 혼합 효과 분석에서 유의하게 나타난 상호작용의 시각화를 위하여 연속변수인 단어 빈도와 고빈도 음절 이웃 수를 다음과 같이 분류하였다. 먼저, 상용 로그로 변환된 백만 어절당 출현 빈도 값(log-transformed occurrences per million)이 0.5 이상일 때 고빈도 단어로, 0.5 미만일 때 저빈도 단어로 데이터를 분류하였으며, 고빈도 음절 이웃 수의 존재 여부에 따라 존재하지 않는 경우(HFSN = 0)와 하나 이상 존재하는 경우(HFSN ≥ 1)로 데이터를 분류하였다. 명사 단어의 철자형 음절 빈도 모형 시각화 결과는 Figure 1에, 음운형 음절 빈도 모형 시각화 결과는 Figure 2에 제시된 바와 같다.

용언 원형에 대한 선형 혼합 효과 모형(Table 7 참고)에서는 위계적 회귀 분석과 마찬가지로 모든 철자형 음절 모형에서 음절 빈도의 유의한 촉진 효과를 보였으며(Table 7 우측 참고. 길이 무관 타입 빈도: β = −.014, SE = .005, t = −2.68, p = .007; 길이 무관 토큰 빈도: β = −.011, SE = .004, t = −2.53, p = .011; 길이 제한 타입 빈도: β = −.022, SE = .005, t = −4.54, p < .001; 길이 제한 토큰 빈도: β = −.015, SE = .004, t = −3.64, p < .001), 이는 음운형 음절의 길이 제한 타입 및 토큰 빈도 모형에서도 마찬가지였다(Table 7 Model 3, 4 참고. 길이 제한 타입 빈

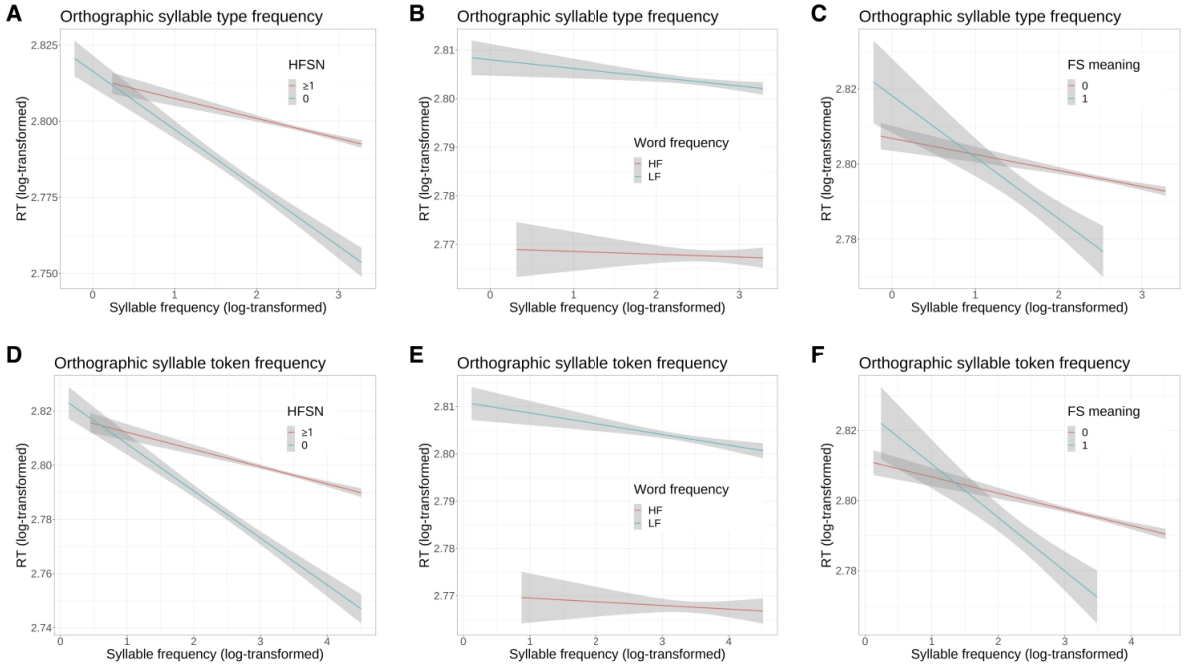

Figure 1. Model estimates of the effect of orthographic syllable type (A, B, C) and token (D, E, F) frequency, by the number of higher frequency syllabic neighbors (A, D), word frequency (B, E), and first syllable meaning (C, F) on reaction time data (log-transformed, ms) for simple nouns.

Note. RT = reaction time; HF = high word frequency (above 0.5 log-transformed occurrences per million); LF = low word frequency (below 0.5 log-transformed occurrences per million); HFSN = number of higher frequency orthographic syllable neighbors (0: first syllable with no higher frequency syllabic neighbor, ≥: first syllable with one or more higher frequency syllabic neighbors; FS meaning = first syllable meaning (0: first syllable with no meaning, 1: first syllable with one or more meanings).

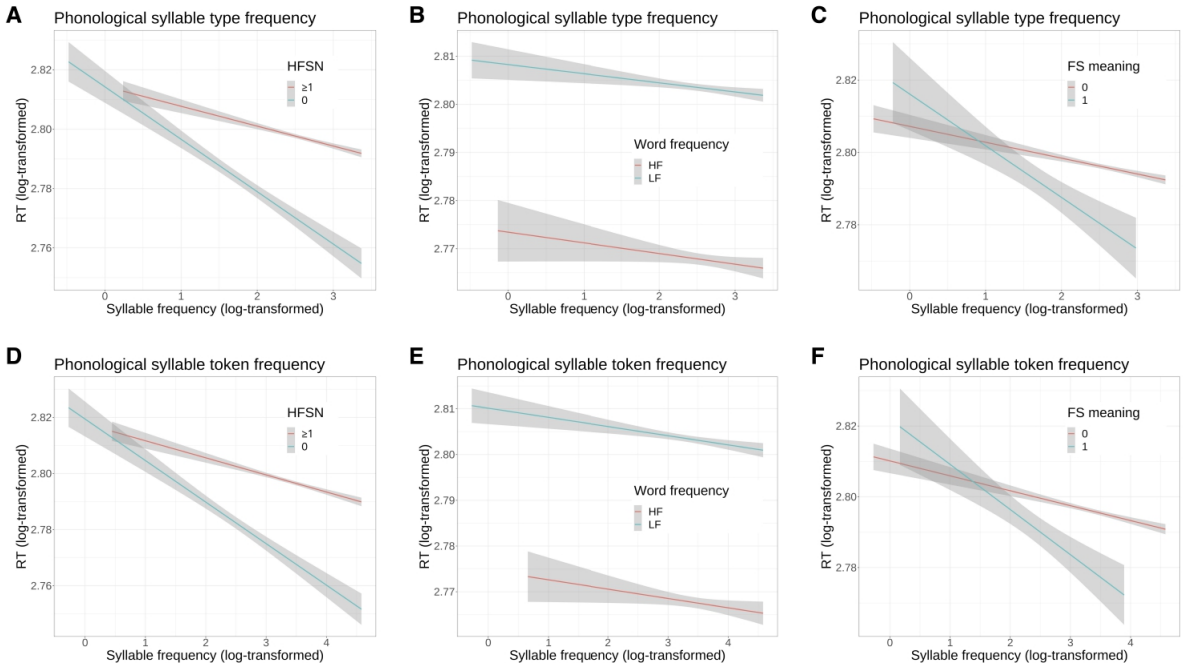

Figure 2. Model estimates of the effect of phonological syllable type (A, B, C) and token (D, E, F) frequency, by the number of higher frequency syllabic neighbors (A, D), word frequency (B, E), and first syllable meaning (C, F) on reaction time data (log-transformed, ms) for simple nouns.
Note. RT = reaction time; HF = high word frequency (above 0.5 log-transformed occurrences per million); LF = low word frequency (below 0.5 log-transformed occurrences per million); HFSN = number of higher frequency orthographic syllable neighbors (0: first syllable with no higher frequency syllabic neighbor, ≥: first syllable with one or more higher frequency syllabic neighbors; FS meaning = first syllable meaning (0: first syllable with no meaning, 1: first syllable with one or more meanings).

도: β = −.018, SE = .005, t = −3.82, p < .001; 길이 제한 토큰 빈도: β = −.009, SE = .004, t = −2.34, p = .019).

다음으로, 철자형 및 음운형 음절의 길이 제한 타입 및 토큰 빈도 모형에서 명사 단어와 마찬가지로 고빈도 음절 이웃 수의 억제 효과가 모두 유의하였다(Table 7 Model 3, 4 참고). 이와 더불어, 고빈도 음절 이웃 수와 음절 빈도 간의 상호작용 역시 유의하게 나타났으며(Table 7 Model 3, 4 참고. 철자형 길이 제한 타입 빈도: β = −.001, SE = 0, t = −4.31, p < .001; 철자형 음절 길이 제한 토큰 빈도: β = −.001, SE = 0, t = −3.15, p = .002; 음운형 음절 길이 제한 타입 빈도: β = −.001, SE = 0, t = −3.93, p < .001; 음운형 음절 길이 제한 토큰 빈도: β = 0, SE = 0, t = −2.35, p = .019), 명사 단어와 마찬가지로 고빈도 음절 이웃 수가 적을수록 음절 빈도 효과가 크게 나타났다.

마지막으로, 음운형 음절의 길이 제한 타입 및 토큰 빈도 모형(Table 7 우측의 Model 3, 4 참고)에서 첫음절 음운 변동 여부의 촉진적 주효과가 발견되었다. 또한, 음절 빈도와의 상호작용이 유의하게 나타났으며(길이 제한 타입 빈도: β = .012, SE = .005, t = 2.65, p = .008; 음절 길이 제한 토큰 빈도: β = .01, SE = .004, t = 2.26, p = .024), 첫음절의 음운 변동이 없을 때 음절 빈도의 촉진적 효과가 크게 나타났다. 그러나, 위계적 회귀 분석 결과와 달리 철자형 및 음운형 음절의 길이 무관 타입 및 토큰 빈도 모형(Table 7 Model 1, 2 참고)에서 음절 빈도 효과는 유의하지 않았으며, 이 외의 상호작용 역시 보이지 않았다.

용언 원형과 달리, 명사 어절 및 용언 굴절형에 대한 선형 혼합 효과 모형(각 Table 8 및 9 참고)에서는 위계적 회귀 분석 결과와 유사하게 음절 빈도의 주효과가 유의하게 나타났다. 명사 어절의 경우 음절 빈도의 촉진적 주효과가 철자형 음절(Table 8 좌측 참고. 길이 무관 타입 빈도: β = −.017, SE = .006, t = −2.95, p = .003; 길이 무관 토큰 빈도: β = −.015, SE = .005, t = −3.09, p = .002; 길이 제한 타입 빈도: β = −.018, SE = .006, t = −3.1, p = .002; 길이 제한 토큰 빈도: β = −.018, SE = .005, t = −3.8, p < .001)과 음운형 음절(Table 8 우측 참고, 길이 무관 타입 빈도: β = −.016, SE = .005, t = −3.08, p = .002; 길이 무관 토큰 빈도: β = −.013, SE = .0, t = −3.03, p = .003; 길이 제한 타입 빈도: β = −.018, SE = .005, t = −3.26, p = .001; 길이 제한 토큰 빈도: β =

Table 7. Results of linear mixed-effect model of reaction time data for simple verbs

Fixed effects	Orthographic syllable			Phonological syllable		
	β	*SE*	*t*-value	*β*	*SE*	*t*-value
Model 1. Type Frequency						
(Intercept)	2.82	.012	228.36***	2.8	.012	242.98***
Word length in syllable	−.003	.002	−1.57	−.006	.002	−2.47*
Word frequency	−.019	.012	−1.64	−.029	.011	−2.56*
First syllable meaning	−.001	.019	−0.05	−.01	.019	−0.51
Phonological change	−.014	.012	−1.13	−.023	.012	−1.92†
HFSN	.002	.001	1.64	.001	.001	1.2
Syllable Frequency	−.014	.005	−2.68**	−.006	.005	−1.19
Word frequency * Syllable Frequency	.004	.005	0.84	.007	.004	1.54
First syllable meaning * Syllable Frequency	−.011	.011	−0.96	−.002	.01	−0.15
Phonological change * Syllable Frequency	.003	.006	0.54	.009	.005	1.78†
HFSN * Syllable Frequency	0	0	−1.34	0	0	−1.06
Model 2. Token Frequency						
(Intercept)	2.82	.014	196.24***	2.81	.014	206.84***
Word length in syllable	−.004	.002	−2.04*	−.006	.002	−2.84**
Word frequency	−.024	.013	−1.77†	−.031	.013	−2.32*
First syllable meaning	.002	.022	0.08	−.005	.022	−0.24
Phonological change	−.01	.015	−0.7	−.024	.014	−1.72†
HFSN	.001	.001	1.6	.001	.001	1.48
Syllable Frequency	−.011	.004	−2.53*	−.006	.004	−1.38
Word frequency * Syllable Frequency	.004	.004	1.16	.006	.004	1.5
First syllable meaning * Syllable Frequency	−.008	.009	−0.89	−.002	.008	−0.32
Phonological change * Syllable Frequency	.001	.005	0.29	.007	.004	1.64
HFSN * Syllable Frequency	0	0	−1.27	0	0	−1.27
Model 3. Type Frequency (fixed-length)						
(Intercept)	2.8	.008	339.92***	2.8	.008	337.51***
Word length in syllable	.003	.002	1.21	.001	.003	0.29
Word frequency	−.005	.007	−0.66	−.009	.008	−1.21
First syllable meaning	−.008	.013	−0.6	−.017	.014	−1.23
Phonological change	−.013	.008	−1.63	−.023	.008	−2.72**
HFSN	.002	0	5.06***	.002	0	4.44***
Syllable Frequency	−.022	.005	−4.54***	−.018	.005	−3.82***
Word frequency * Syllable Frequency	−.003	.004	−0.76	−.001	.004	−0.25
First syllable meaning * Syllable Frequency	−.011	.011	−0.94	.002	.01	0.21
Phonological change * Syllable Frequency	.004	.005	0.81	.012	.005	2.65**
HFSN * Syllable Frequency	−.001	0	−4.31***	−.001	0	−3.93***
Model 4. Token Frequency (fixed-length)						
(Intercept)	2.82	.012	231.36***	2.81	.012	235.57***
Word length in syllable	−.005	.002	−2.9**	−.006	.002	−3.27**
Word frequency	−.028	.011	−2.61**	−.034	.011	−3.07**
First syllable meaning	−.005	.017	−0.28	−.002	.017	−0.13
Phonological change	−.014	.012	−1.17	−.028	.012	−2.33*
HFSN	.002	.001	3.77***	.001	.001	2.82**
Syllable Frequency	−.015	.004	−3.64***	−.009	.004	−2.34*
Word frequency * Syllable Frequency	.008	.004	2.25*	.009	.004	2.45*
First syllable meaning * Syllable Frequency	−.007	.008	−0.86	−.005	.008	−0.63
Phonological change * Syllable Frequency	.003	.005	0.67	.01	.004	2.26*
HFSN * Syllable Frequency	−.001	0	−3.15**	0	0	−2.35*

Note. All frequencies were transformed to per million values and statistical analyses were conducted on log-transformed values. *HFSN* = higher frequency syllabic neighbors.
† $p < .1$ * $p < .05$ ** $p < .01$ *** $p < .001$

Table 8. Results of linear mixed-effect model of reaction time data for inflected nouns

Fixed effects	Orthographic syllable			Phonological syllable		
	β	*SE*	*t*-value	*β*	*SE*	*t*-value
Model 1. Type Frequency						
(Intercept)	2.76	.015	184.5***	2.74	.014	198.3***
Word length in syllable	0	.003	0.11	.002	.003	0.61
Word frequency	−.031	.009	−3.45***	−.028	.01	−2.96**
First syllable meaning	.022	.02	1.09	.018	.019	0.93
Phonological change	−.009	.018	−0.48	.02	.019	1.03
HFSN	−.002	.002	−0.89	0	.002	−0.17
Syllable Frequency	−.017	.006	−2.95**	−.016	.005	−3.08**
Word frequency * Syllable Frequency	.006	.004	1.72†	.006	.004	1.5
First syllable meaning * Syllable Frequency	−.013	.008	−1.62	−.011	.008	−1.42
Phonological change * Syllable Frequency	.009	.008	1.15	−.002	.007	−0.34
HFSN * Syllable Frequency	.001	.001	0.96	0	.001	0.31
Model 2. Token Frequency						
(Intercept)	2.76	.016	175.65***	2.75	.014	193.37***
Word length in syllable	.001	.003	0.23	.001	.003	0.42
Word frequency	−.036	.011	−3.3**	−.028	.01	−2.68**
First syllable meaning	.022	.022	1	.023	.022	1.09
Phonological change	−.011	.024	−0.47	.01	.021	0.47
HFSN	−.001	.002	−0.9	−.001	.001	−0.89
Syllable Frequency	−.015	.005	−3.09**	−.013	.004	−3.03**
Word frequency * Syllable Frequency	.006	.003	1.92†	.004	.003	1.29
First syllable meaning * Syllable Frequency	−.009	.007	−1.32	−.008	.006	−1.37
Phonological change * Syllable Frequency	.008	.007	1.02	.001	.006	0.18
HFSN * Syllable Frequency	0	0	0.96	0	0	1.02
Model 3. Type Frequency (fixed-length)						
(Intercept)	2.74	.013	215.36***	2.73	.012	224.97***
Word length in syllable	.002	.003	0.48	.003	.003	1.02
Word frequency	−.025	.007	−3.65***	−.023	.007	−3.1**
First syllable meaning	.007	.015	0.47	.004	.015	0.3
Phonological change	−.005	.013	−0.36	.014	.014	0.98
HFSN	−.001	.001	−1.28	−.001	.001	−0.75
Syllable Frequency	−.018	.006	−3.1**	−.018	.005	−3.26**
Word frequency * Syllable Frequency	.005	.004	1.38	.004	.004	1.15
First syllable meaning * Syllable Frequency	−.01	.008	−1.2	−.007	.007	−0.96
Phonological change * Syllable Frequency	.009	.007	1.24	−.001	.007	−0.1
HFSN * Syllable Frequency	.001	0	1.36	0	0	0.98
Model 4. Token Frequency (fixed-length)						
(Intercept)	2.77	.015	180.89***	2.77	.014	192.08***
Word length in syllable	−.002	.003	−0.86	−.002	.003	−0.62
Word frequency	−.032	.009	−3.5***	−.028	.009	−3.11**
First syllable meaning	.029	.019	1.54	.025	.019	1.36
Phonological change	−.028	.021	−1.34	.003	.018	0.18
HFSN	−.001	.001	−0.87	−.001	.001	−1.09
Syllable Frequency	−.018	.005	−3.8***	−.017	.004	−3.89***
Word frequency * Syllable Frequency	.006	.003	1.96†	.005	.003	1.61
First syllable meaning * Syllable Frequency	−.013	.007	−2.01*	−.011	.006	−1.8†
Phonological change * Syllable Frequency	.014	.008	1.9†	.003	.006	0.52
HFSN * Syllable Frequency	0	0	1.02	0	0	1.36

Note. All frequencies were transformed to per million values and statistical analyses were conducted on log-transformed values. *HFSN* = higher frequency syllabic neighbors.
† $p < .1$ * $p < .05$ ** $p < .01$ *** $p < .001$

Table 9. Results of linear mixed-effect model of reaction time data for inflected verbs

Fixed effects	Orthographic syllable			Phonological syllable		
	β	*SE*	*t*-value	*β*	*SE*	*t*-value
Model 1. Type Frequency						
(Intercept)	2.76	.014	203.05***	2.77	.013	215.79***
Word length in syllable	.002	.003	0.53	-.001	.003	-0.4
Word frequency	.002	.008	0.23	-.013	.009	-1.5
First syllable meaning	.006	.017	0.36	.011	.018	0.61
Phonological change	.004	.018	0.22	.034	.02	1.71†
HFSN	-.001	.002	-0.44	-.005	.002	-2.22*
Syllable Frequency	-.02	.006	-3.71***	-.017	.005	-3.42***
Word frequency * Syllable Frequency	-.004	.003	-1.05	.001	.003	0.26
First syllable meaning * Syllable Frequency	-.001	.007	-0.12	-.002	.007	-0.26
Phonological change * Syllable Frequency	-.002	.008	-0.24	-.008	.008	-1.05
HFSN * Syllable Frequency	0	.001	0.51	.001	.001	2.22*
Model 2. Token Frequency						
(Intercept)	2.78	.016	170.69***	2.78	.015	189.33***
Word length in syllable	.002	.003	0.75	.001	.003	0.22
Word frequency	-.011	.01	-1.09	-.023	.01	-2.21*
First syllable meaning	0	.02	0.02	.011	.021	0.51
Phonological change	.005	.022	0.22	.036	.024	1.49
HFSN	.001	.002	0.39	-.002	.002	-1.27
Syllable Frequency	-.02	.005	-3.83***	-.019	.005	-3.91***
Word frequency * Syllable Frequency	.002	.003	0.55	.004	.003	1.38
First syllable meaning * Syllable Frequency	.002	.006	0.3	-.001	.006	-0.17
Phonological change * Syllable Frequency	.001	.007	0.08	-.006	.007	-0.9
HFSN * Syllable Frequency	0	.001	-0.28	.001	0	1.3
Model 3. Type Frequency (fixed-length)						
(Intercept)	2.73	.012	227.44***	2.74	.012	227.57***
Word length in syllable	.009	.004	2.45*	.004	.004	1.13
Word frequency	-.005	.006	-0.82	-.015	.006	-2.39*
First syllable meaning	-.002	.012	-0.13	.003	.013	0.2
Phonological change	.01	.012	0.84	.031	.015	2.14*
HFSN	.001	.001	0.48	-.001	.001	-1.24
Syllable Frequency	-.024	.006	-4.23***	-.021	.005	-3.97***
Word frequency * Syllable Frequency	-.001	.003	-0.25	.003	.003	0.91
First syllable meaning * Syllable Frequency	.002	.007	0.36	.001	.007	0.19
Phonological change * Syllable Frequency	-.007	.008	-0.9	-.009	.007	-1.25
HFSN * Syllable Frequency	0	.001	-0.28	.001	0	1.28
Model 4. Token Frequency (fixed-length)						
(Intercept)	2.77	.014	195.1***	2.77	.013	210.94***
Word length in syllable	-.001	.003	-0.26	-.002	.003	-0.74
Word frequency	-.015	.009	-1.81†	-.022	.008	-2.63**
First syllable meaning	.005	.016	0.33	.012	.017	0.72
Phonological change	.009	.019	0.47	.034	.021	1.62
HFSN	.003	.002	1.58	0	.002	0.06
Syllable Frequency	-.019	.005	-4.01***	-.017	.004	-3.81***
Word frequency * Syllable Frequency	.004	.003	1.29	.005	.003	1.69†
First syllable meaning * Syllable Frequency	0	.006	0.07	-.002	.006	-0.31
Phonological change * Syllable Frequency	-.001	.007	-0.09	-.006	.007	-0.93
HFSN * Syllable Frequency	-.001	.001	-1.43	0	0	-0.01

Note. All frequencies were transformed to per million values and statistical analyses were conducted on log-transformed values. *HFSN* = higher frequency syllabic neighbors.
† $p < .1$ * $p < .05$ ** $p < .01$ *** $p < .001$

−.017, SE = .004, t = −3.89, p < .001) 모두에서 유의하게 나타났다.

용언 굴절형 역시 촉진적 음절 빈도 효과를 보였으며, 이는 철자형 음절 모형(Table 9 좌측 참고, 길이 무관 타입 빈도: β = −.02, SE = .006, t = −3.71, p < .001; 길이 무관 토큰 빈도: β = −.02, SE = .005, t = −3.83, p < .001; 길이 제한 타입 빈도: β = −.024, SE = .006, t = −4.23, p < .001; 길이 제한 토큰 빈도: β = −.019, SE = .005, t = −4.01, p < .001)과 음운형 음절 모형(Table 9 우측 참고. 길이 무관 타입 빈도: β = −.017, SE = .005, t = −3.42, p < .001; 길이 무관 토큰 빈도: β = −.019, SE = .005, t = −3.91, p < .001; 길이 제한 타입 빈도: β = −.021, SE = .005, t = −3.97, p < .001; 길이 제한 토큰 빈도: β = −.017, SE = .004, t = −3.81, p < .001) 모두에서 유의하였다.

음절 빈도와 어휘 및 첫음절 속성 간의 상호작용은 명사 어절의 철자형 음절 길이 제한 토큰 빈도 모형(Table 8 Model 4 참고)에서 나타난 첫음절 의미 유무와의 상호작용(β = −.013, SE = .007, t = −2.01, p = .047), 용언 굴절형의 음운형 음절 길이 무관 타입 빈도 모형(Table 9 Model 1 참고)에서 나타난 고빈도 음절 이웃 수와의 상호작용(β = −.013, SE = .007, t = −2.01, p = .047)이 유의하였으며, 명사 단어에서와 마찬가지로 각각 첫음절이 의미를 지닐 경우, 혹은 고빈도 음절 이웃 수가 적을수록 음절 빈도의 촉진적 효과가 크게 나타나는 결과 패턴을 보였다. 그 외의 주효과로는 단어 빈도가 명사 어절의 모든 모형에서 유의한 촉진 효과를 보였으며(Table 8 참고), 용언 굴절형에서는 철자형 음절의 길이 제한 토큰 빈도 모형과 음운형 음절의 길이 무관 타입 및 토큰 빈도, 길이 제한 토큰 빈도 모형에서 유의한 촉진 효과를 보였다(Table 9 우측의 Model 1, 2, 4 참고). 용언 굴절형에서는 이외에도 단어 길이가 철자형 음절의 길이 제한 타입 빈도 모형에서 유의하였으며(Table 9 좌측의 Model 4 참고), 고빈도 음절 이웃 수가 음운형 음절의 길이 무관 타입 빈도 모형에서, 첫음절의 음운 변동 여부가 음운형 음절의 길이 제한 타입 빈도 모형에서 유의하게 나타났다(각각 Table 9 우측의 Model 1과 3 참고).

종합하면, 어휘 변인 및 첫음절 속성 변인과 음절 빈도 간의 상호작용을 고려한 선형 혼합 효과 모형을 사용하여 재분석한 결과, 명사 단어에서는 철자형 음절과 음운형 음절 모두 길이 제한 음절 빈도 모형에서는 음절 빈도의 촉진적 주효과와 더불어 음절 빈도와 첫음절 의미 유무, 고빈도 음절 이웃 수 및 단어 빈도(타입 빈도 제외) 간의 상호작용이 유의하게 나타났다. 길이 무관 음절 빈도에서는 음절 빈도의 주효과는 유의하지 않았으나, 음절 빈도와 단어 빈도, 고빈도 음절 이웃 수 및 음절 빈도 간의 상호작용이 모두 유의하게 나타났으며, 첫음절의 의미가 존재하거나, 단어 빈도가 고빈도일 때, 혹은 고빈도 음절 이웃 수가 없을 때 음절 빈도의 촉진적 효과가 뚜렷하게 나타났다. 용언 원형에서는 음운형 음절의 길이 무관 빈도 모형을 제외한 모든 모형에서 위계적 회귀 분석과 마찬가지로 음절 빈도의 주효과가 촉진적으로 나타났으며, 특히 길이 제한 빈도 모형에서는 명사 단어에서 보인 단어 빈도, 고빈도 음절 이웃 수와 음절 빈도 간의 상호작용 또한 유의하였다. 이와 더불어 첫음절의 의미 변동 여부와 음절 빈도의 상호작용이 발견되었는데, 첫음절의 의미가 존재하지 않을 때 음절 빈도의 촉진 효과가 크게 나타났다. 명사 어절과 용언 굴절형의 경우, 위계적 회귀 분석 결과와 일치하는 결과를 보였으며 철자형 및 음운형 음절, 타입 및 토큰, 길이 무관 및 제한 빈도 모두에서 음절 빈도의 촉진 효과가 유의하게 나타났다.

논 의

본 연구는 한국어 시각 어휘 재인에서 나타나는 음절 빈도 효과를 확인하고, 단어와 어절, 음운과 철자 음절에 따라 보고된 혼재된 결과를 기존 데이터의 재분석을 통해 조사하고자 수행되었다. 이를 위하여 첫음절 빈도를 철자, 음운에 따라 각 네 가지(타입 vs. 토큰; 길이 무관 vs. 길이 제한) 계산 방식으로 나누어 독립적인 위계적 회귀 분석을 수행하였다. 또한, 선형 혼합 효과 모형을 통해 네 어휘 범주에서의 음절 빈도 효과와 각 변인과의 상호작용을 고려한 분석을 추가적으로 진행하였다. 연구 결과, 단어와 어절 간 상이한 패턴의 음절 빈도 효과가 발견되었다. 명사 어절, 용언 원형 및 굴절형 어절은 위계적 회귀 및 선형 혼합 효과 분석 모두에서 음운형과 철자형, 타입 및 토큰, 길이 무관 혹은 길이 제한 빈도 모두 첫음절 빈도가 증가하면 어절 판단 시간이 감소하는 일관적인 촉진 효과가 발견된 반면, 명사 단어는 선형 혼합 효과 분석에서만 일부 모형에서 촉진적 주효과가 나타났으며, 모든 모형에서 음절 빈도와 다른 변인과의 상호작용이 나타났다.

본 연구는 음운형 첫음절 빈도가 단어 재인 시간에 대한 억제 효과를, 철자형 첫음절 빈도는 어절 재인에 대한 촉진 효과를 보일 것으로 예상하였으나, 분석 결과 명사 어절, 용언 원형 및 용언 굴절형에서 첫음절의 철자형과 음운형에 관계없이 어휘 판단 시간에 대한 촉진 효과가 나타났다. 이는

기존 한국어 명사 단어 대상 연구(e.g., Kwon, 2012), 혹은 스페인어(e.g., Conrad et al., 2008)에서 보고된 첫음절 빈도의 억제 효과와 일치하지 않는 결과이다. 기존에 음절 빈도 효과를 설명하는 대표적인 모형으로는 다중 판독 모형(MROM: multiple read-out model)이 제안되었다(Grainger & Jacobs, 1996). 다중 판독 모형은 상호 활성화 모형(interactive activation model; McClleland & Rummelhart, 1981)[5]을 기반으로 하여 시각 단어 재인을 설명하기 위해 확장한 모형으로, 하위 처리 단계로 자소(grapheme)와 음소(phoneme) 단계를 지니고 있으며 자소와 음소 처리 단계는 철자 심성 어휘집(orthographic lexicon)과 음운 심성 어휘집(phonological lexicon)으로 연결된다. 음절 정보는 음운형 하위 어휘 요인에 속하기 때문에 철자 처리 경로가 아닌 음운 처리 경로를 따라 음운형 심성 어휘집 상에서 후보군을 활성화시키는데, 이때 음절의 빈도가 높을수록 활성화되는 후보군의 수가 늘어나 후보군 간의 외측 억제(lateral inhibition)가 발생하여 시각 단어 재인이 지연된다. 그러나 스페인어 연구에서도 음운 단위인 음절이 아니라 철자 하위 단위의 영향력이 공존하고 있다는 결과가 보고된 바 있다(Conrad et al., 2009). 알파벳 계통 언어에서 제안된 하위 어휘 요인 중 음절은 음운형 단위에 해당하며(Carreiras et al., 1993; Conrad et al., 2008), 두 철자(bigram)는 철자형 단위에 해당한다(Seidenberg & McClleland, 1989). Conrad 등(2009)에서는 철자 하위 어휘 요인인 두 철자 쌍의 출현 빈도(bigram frequency)와 음절 빈도를 두 개의 실험으로 나누어서 어휘 판단 과제를 시행하였는데, 그 결과 두 철자 빈도의 촉진 효과와 첫음절 빈도의 억제 효과가 모두 보고되었다. 이에 Conrad 등(2010)은 MROM에서 더 나아가 음절 표상 단위를 추가한 MROM-S(multiple read-out model-syllable representation)을 제안하였다. MROM-S는 최하위 계층에 철자 정보 입력기를 상정하며, 물리적 시각 정보가 처리되면 개별 철자의 표상이 이루어지고 철자 표상 이후 낱자가 조합되어 음절로 표상되는 단계로 구성된다. 이후 음절 표상이 이루어지면 상위 심성 어휘집에서 단어 수준의 후보군 활성화가 이루어져 경쟁이 발생하고, 이로 인하여 어휘 재인에서 억제 효과가 나타나게 된다. 기존 MROM이나 상호 활성화 모형과 달리, MROM-S는 철자가 직접적으로 심성 어휘집으로 연결될 수 있다고 설명한다. 그러나 스페인어와 같이 낱자가 순차적으로 나열되는 풀어쓰기 형식의 쓰기 체계(writing system)를 사용함과 동시에 철자-음운 대응이 투명한 언어는 개별 철자의 영향이 심성 어휘집의 후보군을 활성화할만큼 충분히 크지 않기 때문에, 상위 단계인 음절의 영향력이 상대적으로 강하게 나타나는 것으로 볼 수 있다.

한편, 한국어 연구에서는 스페인어와 같은 두 철자가 아니라 철자형 음절 자체가 시각 단어 재인을 촉진시킨다는 결과가 보고되었다(Yi et al., 2005; Bae & Yi, 2010; Choi et al., 2015; Lim et al., 2022). Choi 등(2015)에서는 점화 어휘 판단 과제(primed lexical decision task)를 이용한 두 개의 실험에서 점화-목표 자극 간 철자 및 음운 관련성에 따른 음절 빈도 효과를 조사하였다. 그 결과, 음운 관련 조건에서는 어휘 판단 시간에 대한 억제적 효과가 나타난 반면, 철자 관련 조건에서는 촉진적 효과가 나타났다. Yi 등(2005)에서는 형태 점화 과제(form priming task)를 사용하여 점화 자극과 목표 자극 간의 철자, 음운 음절 관련성에 따른 효과를 조사하였다. 표기 일치 조건(e.g., 선불-선량)은 점화 자극과 목표 자극이 동일한 표기 음절을 공유하였으나 음운 음절이 일치하지 않았으며, 음운 일치 조건(e.g., 설교-선량)은 동일한 음운 음절을 공유하였으나 철자 음절이 일치하지 않는 조건이었다. 철자-음운 일치 조건(e.g.. 선로-선량)은 점화 자극과 목표 자극이 동일한 철자 및 음운 음절을 공유하였으며, 무관련 조건(e.g, 액체-선량)은 어떠한 음절 관련성도 가지지 않았다. 네 조건을 비교한 결과, 음운 정보가 통제된 상황(음운 일치 조건 vs. 표기-음운 일치 조건)에서 철자 정보의 일치 여부가 시각 단어 재인에 영향을 미친 것으로 나타난 반면, 철자 정보가 통제된 상황(표기 일치 조건 vs. 표기-음운 일치 조건)에서는 음운 정보의 일치 여부가 유의한 영향을 보이지 않았다. 이처럼 한국어와 스페인어에서 공통적으로 철자 부호가 시각 재인에 영향을 미치는 것으로 나타났으나, 스페인어는 음절 자체가 아닌 두 철자(bigram)의 영향을, 한국어는 음절의 철자 표상의 영향을 받고 있다는 차이가 존재한다. 이처럼 두 언어 간의 철자 하위 단위의 차이가 발생하는 이유는 쓰기 체계의 차이로 볼 수 있다. 스페인어와 같이 풀어쓰기를 사용하는 알파벳 계통 언어와 달리, 한국어는 낱자에 해당하는 음소가 집약적으로 결합하여 음절을 구성하는 모아쓰기 방식을 사용하기 때문에, 음절 간의 경계가 시각적으로 뚜렷하게 구분되기 때문이다(Zagar, 2015; Kwon, 2020).

이와 관련하여, Bae와 Yi(2010)는 음운 정보보다 철자 정보가 우선시되는 이중경로모형(dual-route cascaded model;

5) 상호 활성화 모형은 시각 및 청각에 기반한 감각 정보 수준(level)과 중간 단계인 철자, 음소 수준, 단어 수준 3단계로 구성된 모형으로, 이 모형에 의하면 상향식 및 하향식 정보처리를 통해 단어 수준에서 후보군을 활성화하여 단어 재인이 이루어진다.

Coltheart et al., 2001)의 어휘경로(lexical route)와 같이 철자 음절이 어휘 항목과 직접 연결되어 있는 경로를 제안한 바 있으며, 한국어 시각 어휘 재인에서 철자 음절이 기본적인 단위로서 기능할 수 있음을 시사하였다. 한국어는 철자-음운 간의 대응(mappings of grapheme-to-phoneme)이 일관적인 언어로 알려져 있으며(Pae et al., 2020), 음운 변동의 영향이 적어 음절의 음운형과 철자형이 일치하는 경우가 많다. 실제로 본 연구에 사용된 자극에서 첫음절의 음운 변동 비율이 높지 않았으며(명사 단어: 9.21%, 명사 어절: 13.54%, 용언 원형: 18.76%, 용언 굴절형: 20.93%), 음운형 빈도와 철자형 빈도 간의 상관이 높게 나타났다(부록 I 참고). 한편, 본 연구에서는 철자형과 음운형 음절 간의 일치 여부가 시각 어휘 재인, 특히 음절 빈도 효과에 미치는 영향을 알아보고자 첫음절의 음운 변동 여부를 위계적 회귀 분석 모형 2단계의 첫음절 속성 변인 및 선형 혼합 효과 모형의 고정 변인으로 투입하였다. 그 결과, 첫음절 음운 변동 여부의 촉진적 주효과가 명사 단어의 철자형 및 음운형 음절의 길이 제한 토큰 빈도 모형과 용언 원형의 음운형 길이 제한 타입 및 토큰 빈도 모형에서 나타났을 뿐, 그 외의 모형에서는 유의하지 않았다. 다만, 용언 원형의 음운형 음절 길이 제한 타입 및 토큰 빈도 모형(Table 7 우측의 Model 3, 4 참고)에서는 첫음절의 음운 변동 여부와 음절 빈도 효과 간의 상호작용 역시 유의하였는데, 첫음절의 음운 변동이 없는 경우 촉진적 음절 빈도 효과가 강하게 나타났다.

종합하면, 한국어는 모아쓰기 체계로 인하여 음절 간의 시각적 경계가 뚜렷하게 나타나고, 따라서 한국어 시각 어휘 재인이 MROM-S에서 제안한 철자 표상 단계에서 직접 심성 어휘집으로 연결되는 경로를 통해 이루어져 첫음절 빈도의 촉진 효과를 보인 것으로 해석할 수 있다.

본 연구의 위계적 회귀 분석 결과에서는 명사 어절, 용언 원형 및 굴절형을 포함한 어절에서 음절 빈도 효과가 유의하게 나타난 반면, 명사 단어에서는 음절 빈도 효과가 유의하지 않았다. 이러한 상이한 효과의 원인은 고빈도 음절 이웃 수와 전체 형태 빈도의 영향으로 유추할 수 있다. 먼저, 단어와 어절의 고빈도 음절 이웃 수가 보인 차이 또한 위계적 회귀 분석에서 명사 단어의 음절 빈도 효과가 유의하지 않은 결과에 기여했을 가능성이 있다. 본 연구에서는 명사 단어에 대한 상관 분석에서 고빈도 음절 이웃 수와 어휘 판단 시간 간의 정적 상관관계가 나타났으며(HFOSN: r = 0.21; HFPSN: r = 0.20; all ps < .001, 부록 I-1 참고), 명사 단어 대상 선형 혼합 효과 분석(Table 6 참고)에서도 모든 음절 빈도에서 억제적 주효과를 보였다. 이는 기존 연구에서 나타난 고빈도 음절 이웃이 다른 후보군에 비해 더 강하게 활성화되어 목표 어휘의 시각 재인이 지연되는 결과와 일치하는 바이다(Perea & Carreiras, 1998). 실제로 고빈도 음절 이웃 수를 통제한 스페인어 연구에서는 명사 단어에서 첫음절 타입 빈도의 촉진 효과를 보고한 바 있으며(Conrad et al., 2008), 본 연구의 선형 혼합 효과 분석에서는 명사 단어의 모든 모형에서 고빈도 음절 이웃 수와 음절 빈도 간의 상호작용이 유의하였으며, 고빈도 음절 이웃 수가 적을수록 뚜렷한 음절 빈도의 촉진 효과가 발견되었다(Table 6 및 Figure 1A, 1D, 2A, 2D 참고). 따라서, 본 연구 결과는 명사 단어의 고빈도 음절 이웃 수가 다른 세 어휘 범주에 비해 높은 점(Table 1 참고)과 더불어 특히 명사 단어에서 고빈도 음절 이웃 수가 중요한 역할을 한다는 점을 시사한다.

단어와 어절에서 나타난 고빈도 음절 이웃 수의 차이는 한국어 어절의 구성 방식에 의한 것으로 볼 수 있다. 한국어 어절은 내용 형태소(어근 혹은 어간)에 기능 형태소(조사 혹은 어미)가 붙어서 형성되는 활용형으로, 조사나 어미가 붙으면서 시각적 형태의 변화를 수반하는 어휘가 많다. 특히, 용언의 경우 어간 뒤에 붙을 수 있는 어미의 가짓수가 많아, 용언 활용형의 어간이 변화할 가능성이 높기 때문에 화자가 원 형태의 어간을 추측하기 어렵다는 특성이 있다(Cho et al., 2018). 한국어 어절의 이러한 구성 방식은 동일한 어근 혹은 어간을 지니더라도 후에 붙는 조사나 어미에 따라 별개의 후보군으로 여겨질 수 있다. 예를 들어, '가격'이라는 단어에 주격 조사인 '은'이 결합된 어절 '가격은'과 목적격 조사인 '을'이 결합된 어절 '가격을'은 모두 본 연구의 빈도 산출에 사용된 세종 말뭉치(Kang & Kim, 2009)에 별개의 어휘로 취급된다. 이와 같이 어절은 단어와 달리 단일 어근에서 여러 어절이 생성될 수 있기 때문에 동일한 첫음절을 사용한다고 해도 단어에 비해 어절의 음절 이웃 수가 증가하게 된다. 음절 이웃의 누적 빈도가 동일할 때 음절 이웃의 후보군이 많다면, 각 후보군의 빈도가 분산되어 고빈도 이웃의 수가 감소하는 결과로 이어질 수 있다. 이는 실제로 명사 및 용언의 원형 대비 활용형 어절에서 타입 빈도가 높게 나타난 점에서 유추할 수 있다(Table 1 참고). 이러한 두 유형 간의 타입 빈도 차이는 한국어 어절의 구성 방식이 반영된 결과로 볼 수 있으며, 어절의 생산성(productivity)으로 인해 고빈도 이웃의 수가 감소한 것이 음절 빈도 효과에 영향을 미친 것으로 해석할 수 있다.

두 번째로, 전체 형태 빈도는 단어와 어절 두 어휘 형태에서 각각 단어 빈도, 어절 빈도라 이르며, 출현 빈도가 증가할수록 어휘 판단 시간이 감소하는 촉진 효과를 보이는 대표

적인 어휘 요인이다(e.g., 단어: Yi et al., 2017; 어절: Kim & Nam, 2018). 본 연구에서 역시 상관 분석 결과에서 네 어휘 범주 모두 전체 형태 빈도와 어휘 판단 시간 간의 유의한 부적 상관관계를 보인 바 있다(명사 단어: r = −0.34; 명사 어절: r = −0.45; 용언 원형 어절: r = −0.13; 용언 굴절형 어절: r = −0.25; all ps 〈 .001, 부록 I 참고). 그러나 시각 어휘 재인 과정에서 전체 형태 빈도가 음절 빈도 효과에 미치는 영향은 단어와 어절이라는 어휘 범주에 따라 혼재된 결과가 보고되었다. 한국어 어절을 대상으로 한 연구(Kim & Nam, 2018; Kim et al., 2020)에서는 모두 어절의 출현 빈도를 통제하지 않고 다중 회귀 분석에 투입하였으나, 첫음절 타입 및 빈도의 유의한 촉진 효과를 보고하였다(타입 빈도: Kim et al., 2020; 토큰 빈도: Kim & Nam, 2018).

그러나 단어를 대상으로 한 MROM(Grainger & Jacobs, 1996)과 이를 계승한 MROM-S(Conrad et al., 2010)에 따르면, 단어 빈도가 낮을수록 어휘 판단에 필요한 활성화 역치가 높아지므로 이웃 빈도 효과가 고빈도가 아닌 저빈도 단어에서 나타날 것을 예상한다. 이에 대한 경험적 증거는 스페인어 연구(Conrad et al., 2009; Conrad et al., 2010)와 한국어 연구(Kwon, 2012)에서 저빈도 단어에서만 나타난 음절 빈도 효과로 제시된 바 있다. 다만 본 연구의 단어 빈도와 음절 빈도 변인 간의 상호작용을 고려한 추가 분석에서는 저빈도 명사 단어에서 음절 빈도의 촉진적 효과가 뚜렷하게 나타났으며(Table 6와 Figure 1B, 1E, 2B, 2E 참고), 특히 길이 제한 음절 빈도에서는 음절 빈도의 주효과 또한 관찰되어, 저빈도 단어뿐 아니라 고빈도 단어에서도 이러한 촉진적 효과가 유의하게 나타났다. 한편, 한국어 명사 단어를 대상으로 첫음절의 타입, 토큰 빈도의 효과를 비교한 Jin 등(2018)에서는 단어 빈도를 저빈도로 통제하였으나, 타입과 토큰 빈도 모두 유의한 효과를 보이지 않았다. 이에 대하여 저자들은 적은 수의 실험 자극 혹은 실험 참가자로 인한 통계적 검증력의 문제를 그 원인으로 추측하였으나, 본 연구에서는 20,959개의 명사 단어에 대한 대규모 어휘 판단 데이터를 대상으로 위계적 다중 회귀 기법을 이용한 재분석을 진행한 결과 유의한 음절 빈도 효과를 관찰하는 데 실패하였다. 다만, 선형 혼합 효과 모형을 이용한 추가 분석 결과 길이 제한 음절 빈도에서는 촉진적 주효과가, 길이 무관 음절 빈도에서는 단어 빈도와의 상호작용이 관찰된 바, 향후 연구에서는 요인 설계 실험을 통하여 명사 단어 대상 단어 빈도와 음절 빈도 간 변인을 직접적으로 조작하여 조사할 필요성을 제기한다.

마지막으로, 첫음절이 지닌 의미 유무 역시 명사 단어가 다른 세 어휘 범주와 상이한 음절 빈도 효과의 패턴을 보인 까닭으로 유추할 수 있다. 본 연구의 선형 혼합 효과 분석 결과, 길이 제한 타입 빈도 모형을 제외한 명사 단어의 모든 음절 빈도 모형에서 첫음절의 의미 유무가 주효과를 보였으며, 첫음절의 의미가 있는 경우 그렇지 않은 경우보다 반응 시간이 빨라지는 촉진적 효과를 보였다(Table 6 참고). 이는 첫음절의 한자 형태소를 공유하는 의미 이웃이 명사 단어의 시각 재인에 영향을 미친다는 결과를 보고한 Kwon과 Nam(2011)의 경험적 증거를 일부 지지하는 결과이다. 그러나 길이 제한 타입 및 토큰 빈도 모형에서 철자형 및 음운형 음절 모두에서 촉진적 음절 빈도 효과를 발견한 본 연구와 달리, Kwon과 Nam(2011)은 음운 음절의 이웃 크기(음운형 음절의 타입 빈도)의 효과를 조사한 실험 1에서 음절 빈도의 억제적 주효과를 보고하였으며, 표기 음절의 이웃 크기(즉, 철자형 음절의 타입 빈도)의 효과를 조사한 실험 2에서는 유의미한 음절 빈도 효과를 발견하지 못하였다. 이와 더불어, 실험 1에서는 음운형 음절의 타입 빈도가 높으면서 한자어 이웃이 큰 경우가 작은 경우에 비해 어휘 판단 시간이 짧아지는 결과를 보인 반면, 실험 2에서는 이와 같은 상호작용이 피험자별 분석에서만 유의하였을 뿐, 사후 분석에서는 유의미한 차이를 보이지 않았다. 이와 달리, 본 연구는 명사 단어 대상 선형 혼합 효과 분석의 모든 모형에서 첫음절 의미 유무와 음절 빈도 간의 상호작용이 유의하였으며(Table 6와 Figure 1C, 1F, 2C, 2F 참고), 첫음절이 의미를 가질 때 촉진적 음절 빈도 효과가 강하게 나타났다. 결론적으로, 기존 연구와 본 연구 결과를 종합하면 한국어 시각 어휘 재인 과정에서 적어도 명사 단어에서는 첫음절의 의미 이웃 유무, 고빈도 음절 이웃 수와 첫음절 자체가 가진 속성과 더불어 단어 빈도라는 어휘 변인이 음절 빈도 효과에 영향을 미친다고 볼 수 있다. 다만, 기존 연구와 달리 대규모 데이터를 활용하여 세밀한 분석 기준을 사용한 본 연구에서는 음절 이웃 크기(타입 빈도), 혹은 음절의 누적 빈도(토큰 빈도)가 크거나 높을 때 모두 일관적으로 촉진적인 방향성을 보였다.

요약하면, 본 연구는 품사, 형태에 따라 상이한 결과가 보고된 한국어 시각 어휘 재인 시의 음절 빈도 효과를 조사하기 위하여 기존 어휘 판단 데이터를 품사와 형태, 빈도 계산 방식에 따라 세분화하여 재분석하였다. 위계적 회귀 분석 결과에서는 명사 어절, 용언 원형 및 용언 굴절형에서 타입과 토큰, 길이 무관과 길이 제한 등의 빈도 계산 방식과 관계없이 일관적인 첫음절 빈도의 촉진 효과가 나타났으며, 선형 혼합 효과 분석 결과에서는 세 어휘 범주뿐만 아니라 명사

단어 역시 길이 제한 음절 빈도 모형에서 첫음절 빈도의 유의한 촉진 효과를 보였다. 이와 같은 촉진적 음절 빈도 효과는 첫음절의 철자 또는 음운 표상과 무관하게 일관된 결과를 보였는데, 이는 한국어의 모아쓰기 체계로 인하여 음절 간의 시각적 경계가 뚜렷하여 철자 음절의 표상을 통한 어휘 접속(lexical access)이 가능한 것으로 유추된다. 한편, 위계적 회귀 모형에서 명사 단어가 네 어휘 범주 중 유일하게 촉진적 음절 빈도 효과를 보이지 않은 점과 더불어, 명사 단어 대상 선형 혼합 효과 분석에서 단어 빈도, 첫음절 의미 유무, 혹은 고빈도 음절 이웃 수와 음절 빈도 간의 상호작용이 일관적으로 나타난 점은 단어와 어절에서 음절이 상이하게 기능할 가능성을 제기한다.

한편, 본 연구에서 고려하지 못한 몇 가지 제한점이 존재한다. 먼저, 첫음절 빈도의 억제 효과를 보고한 기존 시각 단어 재인 연구에서는 앞서 어절에서 나타난 상이한 결과의 원인으로 언급된 고빈도 음절 이웃 수가 통제되었을 때 비로소 유의한 효과가 나타났다(Conrad et al. 2008; 2009; Kwon, 2012). 본 연구에서는 어휘의 범주(명사 단어, 명사 어절, 용언 원형 및 용언 굴절형) 간의 고빈도 이웃 수가 차이를 보인 바 있으며(Table 1 참고), 선형 혼합 효과 분석 결과, 명사 단어뿐 아니라 용언 원형의 길이 제한 음절 빈도 모형 역시 고빈도 음절 이웃 수의 유의한 억제적 효과가 보고된 바, 추후 연구에서는 어절을 대상으로 해당 변인에 대한 적절한 통제가 이루어진 후 요인 설계 연구를 통하여 음절 빈도 효과를 조사할 필요성이 제기된다. 다음으로, 본 연구에서는 어휘 범주 간 동질성 확보를 위하여, 첫음절의 타입 및 토큰 빈도, 고빈도 음절 이웃 수와 같은 음절 관련 수치를 세종 말뭉치(Kim & Kang, 2009) 내의 어휘 범주별 별도의 말뭉치 내에서가 아닌 1500만 어절 규모의 전체 형태 말뭉치를 기준으로 산출하였다. 혹자는 이와 같은 산출 방식에 대하여, 각 어휘 범주의 음절 이웃에 다른 어휘 범주가 포함되는 문제로 인해 단어와 어절 각각에서 음절의 역할이 축소되는 결과를 유발할 수 있다는 가능성을 제기할 수 있다. 예를 들어, “가면”이라는 어휘는 일반명사 단독 말뭉치로 음절 빈도를 산출할 경우, “가구”, “가방” 등과 같이 동일한 어휘 범주만이 음절 이웃에 포함되는 반면, 말뭉치 전체를 기준으로 음절 빈도를 산출할 경우 앞서 동일한 어휘 범주 명사 단어인 “가구”, “가방”뿐 아니라 “가도”, “가서”와 같은 동사 어절까지 음절 이웃에 포함되는 차이가 발생하게 된다. 추후에는 동일한 어휘에 대하여 전체 말뭉치 기준 산출 방식과 품사별 말뭉치 기준 산출 방식 간의 비교를 통해 그 영향력을 살펴보는 연구 또한 필요할 것으로 사료된다. 마지막으로, 본 연구는 명사 단어와 용언 원형을 대상으로 한 데이터 재분석을 위하여 기존의 어휘 판단 메가스터디(Yi et al., 2017)의 총 22,052개(명사 단어: 20,959개, 용언 원형: 1,093개)라는 방대한 수의 어휘를 활용하였다. 그러나 어절을 대상으로 한 연구의 수가 현재까지는 부족한 연유로, 어절 판단 데이터는 두 연구(Kim & Nam, 2018; Kim et al., 2020)에서 추출되었으며, 그 결과 명사 어절과 용언 굴절형은 상대적으로 적은 수의 어휘(명사 어절: 480개, 용언 굴절형: 406개)만이 활용 가능하였다. 추후 연구에서는 축적된 어절 대상 연구를 활용하여 이러한 자극 수의 불균형을 해소할 수 있을 것으로 기대된다.

References

Álvarez, C., Carreiras, M., & Perea, M. (2004). Are syllables phonological units in visual word recognition?. *Language and Cognitive processes*, *19*(3), 427-452.

Baayen, R. H., Feldman, L. B., & Schreuder, R. (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language, 55*(2), 290-313.

Bae, & Yi. (2010). Processing of Orthography and Phonology in Korean Word Recognition. *The Korean Journal of Cognitive and Biological Psychology*, *22*(3), 369-385.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of experimental psychology: General, 133*(2), 283.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., Heisterkamp, S., Van Willigen, B., & Maintainer, R. (2017). Package ‘nlme’. *Linear and nonlinear mixed effects models, version, 3*(1), 274.

Carreiras, M., Álvarez, C. J., & Devega, M. (1993). Syllable frequency and visual word recognition in Spanish. *Journal of memory and language*, *32*(6), 766-780.

Carreiras, M., Perea, M., & Grainger, J. (1997). Effects of the orthographic neighborhood in visual word recognition: Cross-task comparisons. *Journal of experimental psychology: learning, memory, and cognition*, *23*(4), 857.

Chetail, F., & Mathey, S. (2011). Effect of syllabic neighbourhood as a function of syllabic neighbour

length. *Psychonomic bulletin & review, 18*, 951-957.

Cho, J., Noh, S., Yu, H., & Doh, W. (2018). Exploration of the Conjugated Forms in the Korea University Korean Dictionary. *Journal of Korealex (32)*, 95-113.

Choi, W., Lee, C., Kang, J., & Nam, K. (2015). The lexical inhibition of the phonological information in Korean visual word recognition. *The Korean Journal of Cognitive and Biological Psychology, 27*(3), 561-581.

Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. *In Attention and performance VI* (pp. 535-555). Routledge.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological review, 108*(1), 204.

Conrad, M., & Jacobs, A. (2004). Replicating syllable frequency effects in Spanish in German: One more challenge to computational models of visual word recognition. *Language and Cognitive Processes*, *19*(3), 369-390.

Conrad, M., Grainger, J., & Jacobs, A. M. (2007). Phonology as the source of syllable frequency effects in visual word recognition: Evidence from French. *Memory & Cognition*, *35*(5), 974-983.

Conrad, M., Carreiras, M., & Jacobs, A. M. (2008). Contrasting effects of token and type syllable frequency in lexical decision. *Language and Cognitive Processes*, *23*(2), 296-326.

Conrad, M., Carreiras, M., Tamm, S., & Jacobs, A. M. (2009). Syllables and bigrams: orthographic redundancy and syllabic units affect visual word recognition at different processing levels. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(2), 461.

Conrad, M., Tamm, S., Carreiras, M., & Jacobs, A. M. (2010). Simulating syllable frequency effects within an interactive activation framework. *European Journal of Cognitive Psychology, 22*(5), 861-893.

Goh, W. D., Yap, M. J., Lau, M. C., Ng, M. M., & Tan, L. C. (2016). Semantic richness effects in spoken word recognition: A lexical decision and semantic categorization megastudy. *Frontiers in psychology, 7*, 976.

Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: a multiple read-out model. *Psychological review, 103*(3), 518

Hutzler, F., Bergmann, J., Conrad, M., Kronbichler, M., Stenneken, P., & Jacobs, A. M. (2004). Inhibitory effects of first syllable-frequency in lexical decision: an event-related potential study. *Neuroscience Letters*, *372*(3), 179-184.

Jin, R., Lee, H., & Choi, W. (2018). Are they real neighbors?: Null effects of syllabic neighbors in Korean word recognition. *The Korean Journal of Cognitive and Biological Psychology*, *30*(3), 211-223.

Kang, B., & Kim, H. (2009). *Frequency of Korean language: Based on 15 million size Eojeol Sejong syntactic and semantic corpus*. Seoul: Handkook munhwasa.

Kim, J., & Nam, K. (2018). Lexical factors that influence the Korean Eojeol recognition. *The Korean Journal of Cognitive and Biological Psychology*, *30*(4), 373-390.

Kim, S., Koo, M., Kim, J., and Nam, K. (2020). The Research for Language Information Processing of Bilateral Hemispheres on Korean Noun Eojeol: Visual Half-field Study. *The Korean Journal of Cognitive and Biological Psychology, 32*(1), 29-53.

Kim, J., Kang, J., Kim, J., & Nam, K. (2022). Temporal dynamics of form and meaning in morphologically complex word processing: An ERP study on Korean inflected verbs. *Journal of Neurolinguistics, 64*, 101098.

Kwon, Y., & Nam, K. (2011). The Relationship Between Morphological Family Size and Syllabic Neighborhoods Density in Korean Visual Word Recognition. *The Korean Journal of Cognitive and Biological Psychology, 23*(3), 301-319.

Kwon, Y., Nam, K., & Lee, Y. (2012). ERP index of the morphological family size effect during word recognition. *Neuropsychologia, 50*(14), 3385-3391.

Kwon, Y. (2012). The Dissociation of Syllabic Token and Type Frequency Effect in Lexical Decision Task. *The Korean Journal of Cognitive and Biological Psychology, 24*(4), 315-328.

Kwon, Y. (2014). The syllable type and token frequency effect in naming task. *The Korean Journal of Cognitive Science, 25*(2), 91-107.

Kwon, Y., & Lee, Y. (2015). The source of the syllable frequency effect during visual word recognition: event-related brain potential study. *Journal of Language*

Sciences, 22(4), 1-17.

Kwon, Y., Nam, K., & Lee, Y. (2015). The role of orthographic syllable frequency in the syllable frequency effect: Evidence from Korean. *Perceptual and motor skills, 120*(1), 95-109.

Kwon, Y. (2020). The review of syllable frequency effect in Korean visual word recognition. *The Korean Journal of Cognitive and Biological Psychology*, *32*, 291-303.

Lee, C., & Taft, M. (2009). Are onsets and codas important in processing letter position? A comparison of TL effect in English and Korean. *Journal of Memory and Language, 60*, 530-542.

Lee, S., Nam, K., Lee, S., Jeon, H., & Kim, Y. (2019). The Influence of Lexical Factors on Verbal Eojeol Recognition: An ERP Study, *The Journal of Linguistic Science, 91,* 289-314.

Lee, Y., & Kwon, Y. (2012). The Effect of the Individual Differences in Working Memory on Sentence Processing. *Journal of The Korean Data Analysis Society, 14*(2), 825-836.

Lim, C., Baek, H., Kim, T. H., & Choi, W. (2022). Activation of Phonological and Orthographic information during Korean Visual Word Recognition: Evidence from a Meta-analysis and a Priming Study. *The Korean Journal of Cognitive Science, 34*(4), 221-236.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological review, 88*(5), 375

Pae, H. K., Bae, S., & Yi, K. (2020). More than an alphabet: Linguistic features of Korean and their influences on Hangul word recognition. *Written Language & Literacy, 22*(2), 223-246.

Perea, M., & Carreiras, M. (1998). Effects off syllable frequency and syllable neighborhood frequency in visual word recognition. *Journal of Experimental Psychology: Human perception and performance, 24*(1), 134.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological review, 96*(4), 523.

Tae, J., Lee, Y., & Kwon, Y. (2017). The Role of Phonological Information on Visual Word Recognition through Auditory-visual Cross Modal Priming. *Journal of Language Sciences, 24*(1), 175-189.

Wickham H. (2009). *ggplot2: elegant graphics for data analysis*. New York: Springer-Verlag.

Yap, M. J., & Balota, D. A. (2009). Visual word recognition of multisyllabic words. *Journal of Memory and Language, 60*(4), 502-529.

Yi, K., Koo, M., Nam, K., Park, K., Park, T. J., Bae, S., Lee, C., Lee, H., & Cho, J. (2017). The Korean Lexicon Project: A Lexical Decision Study on 30,930 Korean Words and Nonwords. *The Korean Journal of Cognitive and Biological Psychology*, 29(4), 395-410.

Zagar, D. (2015). Hangul: A fascinating writing system. A comment on Kwon, Nam, and Lee (2015). *Perceptual and Motor Skills, 121*(2), 461-464.

어휘 재인에서의 음절 빈도 효과: 한국어 단어와 어절에 대한 회귀 분석 연구

김준우[1], 이솔빈[1], 김상엽[2], 남기춘[1,2]

[1]고려대학교 심리학부, [2]고려대학교 지혜과학연구센터

기존 한국어 시각 어휘 재인 연구에서 음절 빈도 효과는 명사 및 동사와 같은 어휘의 품사, 단어 및 어절 등 어휘의 형태, 나아가 타입 및 토큰 빈도와 같은 음절 빈도의 다양한 계산 방식에 따라 혼재된 결과가 보고되었다. 이에 본 연구에서는 철자형 및 음운형 빈도 각각에 대하여, 타입 및 토큰, 길이 제한 및 길이 무관 빈도 등 네 가지 계산 방식에 따라 첫음절 빈도를 계산하여 기존 한국어 어휘 판단 데이터의 23,173개 단어 및 어절 대상 어휘 판단 데이터를 재분석하였다. 총 8가지 독립적 위계적 회귀 모형에서 용언의 원형 및 굴절형 모두에서 첫음절 빈도의 촉진 효과가 나타난 반면, 명사는 어절에서만 촉진 효과가 발견되었다. 이러한 첫음절 빈도의 촉진 효과는 철자 및 음운형 음절 빈도 모두 빈도 계산 방식과 무관하게 일관적으로 나타났다. 음절 빈도의 촉진적 효과는 한국어의 독특한 모아쓰기 체계의 영향으로 음절 간의 경계가 시각적으로 뚜렷해지고, 이로 인하여 시각 어휘 재인 시 직접적 철자 음절 경로를 통한 어휘 접속이 가능하기 때문으로 여겨진다. 또한, 명사 단어에서의 효과 부재는 단어 빈도 효과와 고빈도 이웃 수의 영향으로 유추된다. 따라서, 본 연구 결과는 음절의 음운형 및 철자형 표상 모두가 중요한 역할을 한다는 점을 시사한다.

주제어: 시각 어휘 재인, 음절 빈도 효과, 위계적 회귀 분석, 철자형 표상, 음운형 표상

부록 Ⅰ-1.

Correlations of variables in simple nouns

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. LenSyll		−0.27***	−0.10***	−0.09***	0.13***	0.59***	0.13***	0.24***	0.14***	0.58***	0.14***	0.22***	−0.12***
2. Full-form Frequency			−0.38***	−0.38***	0.10***	−0.09***	0.11***	0.06***	0.09***	−0.09***	0.10***	0.06***	−0.34***
3. HFOSN				0.94***	0.57***	0.42***	0.54***	0.50***	0.54***	0.40***	0.50***	0.48***	0.21***
4. HFPSN					0.53***	0.40***	0.51***	0.48***	0.58***	0.45***	0.56***	0.53***	0.20***
5. Orthographic Type Frequency						0.83***	0.96***	0.89***	0.93***	0.78***	0.88***	0.83***	−0.04***
6. Orthographic Token Frequency							0.81***	0.87***	0.79***	0.95***	0.75***	0.81***	−0.09***
7. Orthographic Type Frequency (Length-restraint)								0.93***	0.90***	0.76***	0.92***	0.86***	−0.05***
8. Orthographic Token Frequency (Length-restraint)									0.83***	0.82***	0.85***	0.91***	−0.07***
9. Phonological Type Frequency										0.84***	0.96***	0.90***	−0.04***
10. Phonological Token Frequency											0.82***	0.87***	−0.09***
11. Phonological Type Frequency (Length-restraint)												0.94***	−0.05***
12. Phonological Token Frequency (Length-restraint)													−0.07***
13. log RT													

Note. All frequencies were transformed to per million values and statistical analyses were conducted on log-transformed values.
LenSyll = word length in syllable; *HFOSN* = higher frequency orthographic syllabic neighbors; *HFPSN* = higher frequency phonological syllabic neighbors; *log RT* = log-transformed response time (ms).
* $p < .05$, ** $p < .01$, *** $p < .001$.

부록 Ⅰ-2.

Correlations of variables in inflected nouns

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. LenSyll		−0.39***	−0.14**	−0.15**	0.18***	0.37***	0.15**	−0.13**	0.05	0.27***	0.04	−0.21***	0.13**
2. Full-form Frequency			−0.42***	−0.41***	0.10*	−0.00	0.15***	0.27***	0.18***	0.06	0.21***	0.31***	−0.45***
3. HFOSN				0.97***	0.34***	0.28***	0.31***	0.36***	0.30***	0.26***	0.27***	0.31***	0.11*
4. HFPSN					0.32***	0.26***	0.30***	0.35***	0.33***	0.29***	0.32***	0.36***	0.14**
5. Orthographic Type Frequency						0.95***	0.94***	0.90***	0.86***	0.86***	0.80***	0.75***	−0.24***
6. Orthographic Token Frequency							0.89***	0.81***	0.79***	0.89***	0.73***	0.66***	−0.20***
7. Orthographic Type Frequency (Length-restraint)								0.93***	0.85***	0.83***	0.87***	0.80***	−0.27***
8. Orthographic Token Frequency (Length-restraint)									0.82***	0.77***	0.81***	0.87***	−0.33***
9. Phonological Type Frequency										0.95***	0.95***	0.91***	−0.27***
10. Phonological Token Frequency											0.89***	0.83***	−0.23***
11. Phonological Type Frequency (Length-restraint)												0.93***	−0.28***
12. Phonological Token Frequency (Length-restraint)													−0.33***
13. log RT													

Note. All frequencies were transformed to per million values and statistical analyses were conducted on log-transformed values.
LenSyll = word length in syllable; *HFOSN* = higher frequency orthographic syllabic neighbors; *HFPSN* = higher frequency phonological syllabic neighbors; *log RT* = log-transformed response time (ms).
* $p < .05$, ** $p < .01$, *** $p < .001$.

부록 Ⅰ-3.

Correlations of variables in simple verbs

	1	2	3	4	5	6	7	8	9	10	11	12	13
1. LenSyll		−0.35***	−0.15***	−0.18***	0.31***	0.53***	0.26***	0.04	0.32***	0.52***	0.28***	0.06*	−0.12***
2. Full-form Frequency			−0.06	−0.10**	0.07*	−0.06	0.14***	0.22***	0.02	−0.09**	0.08**	0.17***	−0.13***
3. HFOSN				0.85***	0.56***	0.45***	0.52***	0.58***	0.50***	0.40***	0.45***	0.52***	0.05
4. HFPSN					0.41***	0.32***	0.39***	0.46***	0.56***	0.45***	0.52***	0.60***	0.04
5. Orthographic Type Frequency						0.95***	0.95***	0.90***	0.84***	0.82***	0.79***	0.74***	−0.14***
6. Orthographic Token Frequency							0.89***	0.82***	0.82***	0.89***	0.77***	0.70***	−0.18***
7. Orthographic Type Frequency (Length-restraint)								0.94***	0.81***	0.78***	0.84***	0.79***	−0.17***
8. Orthographic Token Frequency (Length-restraint)									0.76***	0.71***	0.78***	0.84***	−0.20***
9. Phonological Type Frequency										0.95***	0.95***	0.90***	−0.11***
10. Phonological Token Frequency											0.90***	0.84***	−0.16***
11. Phonological Type Frequency (Length-restraint)												0.94***	−0.14***
12. Phonological Token Frequency (Length-restraint)													−0.16***
13. log RT													

Note. All frequencies were transformed to per million values and statistical analyses were conducted on log-transformed values.
LenSyll = word length in syllable; *HFOSN* = higher frequency orthographic syllabic neighbors; *HFPSN* = higher frequency phonological syllabic neighbors; *log RT* = log-transformed response time (ms).
* $p < .05$, ** $p < .01$, *** $p < .001$.

부록 Ⅰ-4.

Correlations of variables in inflected verbs

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13
1. LenSyll		−0.35***	−0.07	−0.09	0.17***	0.42***	0.15**	−0.07	0.14**	0.39***	0.12*	−0.08	0.03
2. Full-form Frequency			−0.35***	−0.39***	0.21***	0.06	0.27***	0.35***	0.24***	0.10*	0.26***	0.33***	−0.25***
3. HFOSN				0.84***	0.39***	0.35***	0.35***	0.39***	0.32***	0.29***	0.30***	0.33***	−0.03
4. HFPSN					0.24***	0.21***	0.23***	0.26***	0.38***	0.35***	0.36***	0.40***	−0.06
5. Orthographic Type Frequency						0.94***	0.92***	0.90***	0.81***	0.78***	0.76***	0.73***	−0.37***
6. Orthographic Token Frequency							0.86***	0.81***	0.77***	0.85***	0.72***	0.66***	−0.34***
7. Orthographic Type Frequency (Length-restraint)								0.94***	0.81***	0.77***	0.86***	0.80***	−0.36***
8. Orthographic Token Frequency (Length-restraint)									0.79***	0.71***	0.81***	0.85***	−0.37***
9. Phonological Type Frequency										0.94***	0.94***	0.93***	−0.35***
10. Phonological Token Frequency											0.89***	0.84***	−0.32***
11. Phonological Type Frequency (Length-restraint)												0.95***	−0.35***
12. Phonological Token Frequency (Length-restraint)													−0.35***
13. log RT													

Note. All frequencies were transformed to per million values and statistical analyses were conducted on log-transformed values.
LenSyll = word length in syllable; *HFOSN* = higher frequency orthographic syllabic neighbors; *HFPSN* = higher frequency phonological syllabic neighbors; *log RT* = log-transformed response time (ms).
* $p < .05$, ** $p < .01$, *** $p < .001$.

부록 II-1.

Multicollinearity indices TOL (tolerance) / VIF (variance inflation factor) values in models for orthographic syllables

Model	Predictor	Noun				Verb			
		Simple		Inflected		Simple		Inflected	
		TOL	VIF	TOL	VIF	TOL	VIF	TOL	VIF
Model 1-1	LenSyll	0.68	1.46	0.55	1.82	0.54	1.84	0.66	1.52
	Full-form frequency	0.52	1.91	0.45	2.23	0.74	1.36	0.48	2.08
	HFOSN	0.85	1.18	0.88	1.14	0.84	1.19	0.91	1.10
	First syllable meaning	0.91	1.10	0.82	1.23	0.81	1.23	0.73	1.37
	Phonological change	0.37	2.68	0.49	2.06	0.49	2.03	0.53	1.90
	Orthographic Type Frequency	0.38	2.60	0.53	1.9	0.36	2.80	0.45	2.20
Model 1-2	LenSyll	0.70	1.43	0.57	1.76	0.6	1.67	0.67	1.48
	Full-form Frequency	0.54	1.86	0.45	2.22	0.71	1.41	0.48	2.1
	HFOSN	0.86	1.17	0.94	1.06	0.89	1.13	0.94	1.06
	First syllable meaning	0.93	1.07	0.87	1.15	0.87	1.15	0.85	1.17
	Phonological change	0.41	2.46	0.5	2.01	0.56	1.80	0.54	1.84
	Orthographic Token Frequency	0.42	2.36	0.59	1.68	0.44	2.25	0.53	1.89
Model 1-3	LenSyll	0.31	3.18	0.46	2.15	0.39	2.55	0.48	2.06
	Full-form Frequency	0.58	1.73	0.48	2.10	0.78	1.29	0.52	1.92
	HFOSN	0.84	1.20	0.88	1.13	0.86	1.16	0.90	1.11
	First syllable meaning	0.92	1.09	0.80	1.25	0.83	1.21	0.75	1.33
	Phonological change	0.40	2.50	0.51	1.95	0.53	1.89	0.54	1.84
	Orthographic Type Frequency (length-restraint)	0.28	3.63	0.49	2.03	0.32	3.11	0.41	2.42
Model 1-4	LenSyll	0.64	1.57	0.68	1.48	0.75	1.33	0.79	1.26
	Full-form Frequency	0.57	1.75	0.44	2.29	0.71	1.40	0.46	2.15
	HFOSN	0.88	1.14	0.94	1.07	0.91	1.10	0.94	1.06
	First syllable meaning	0.95	1.06	0.85	1.18	0.88	1.14	0.85	1.18
	Phonological change	0.44	2.28	0.49	2.05	0.54	1.84	0.52	1.91
	Orthographic Token Frequency (length-restraint)	0.45	2.20	0.57	1.76	0.48	2.06	0.52	1.94

Note. All frequencies were transformed to per million values and statistical analyses were conducted on log-transformed values.
LenSyll = word length in syllable; HFOSN = higher frequency orthographic syllabic neighbors

부록 II-2.

Multicollinearity indices TOL (tolerance) / VIF (variance inflation factor) values in models for phonological syllables

Model	Predictor	Noun				Verb			
		Simple		Inflected		Simple		Inflected	
		TOL	VIF	TOL	VIF	TOL	VIF	TOL	VIF
Model 1-1	LenSyll	0.69	1.46	0.6	1.67	0.52	1.92	0.63	1.58
	Full-form frequency	0.53	1.9	0.46	2.18	0.74	1.35	0.43	2.31
	HFPSN	0.88	1.13	0.91	1.09	0.91	1.1	0.94	1.07
	First syllable meaning	0.98	1.02	0.93	1.07	0.98	1.02	0.93	1.07
	Phonological change	0.36	2.75	0.5	2.01	0.46	2.16	0.46	2.18
	Phonological Type Frequency	0.39	2.54	0.63	1.6	0.4	2.47	0.51	1.96
Model 1-2	LenSyll	0.7	1.42	0.6	1.67	0.57	1.74	0.65	1.54
	Full-form Frequency	0.54	1.84	0.45	2.21	0.72	1.39	0.44	2.28
	HFPSN	0.9	1.12	0.96	1.04	0.94	1.07	0.95	1.05
	First syllable meaning	0.99	1.01	0.94	1.07	0.99	1.01	0.94	1.07
	Phonological change	0.4	2.51	0.5	2.01	0.53	1.89	0.48	2.09
	Phonological Token Frequency	0.44	2.3	0.65	1.55	0.48	2.1	0.54	1.86
Model 1-3	LenSyll	0.32	3.08	0.51	1.95	0.39	2.59	0.47	2.14
	Full-form Frequency	0.58	1.72	0.48	2.08	0.77	1.3	0.46	2.16
	HFPSN	0.87	1.14	0.91	1.09	0.92	1.09	0.93	1.08
	First syllable meaning	0.98	1.02	0.92	1.08	0.98	1.02	0.93	1.07
	Phonological change	0.39	2.59	0.52	1.94	0.49	2.03	0.47	2.12
	Phonological Type Frequency (length-restraint)	0.28	3.52	0.6	1.65	0.35	2.82	0.46	2.18
Model 1-4	LenSyll	0.65	1.53	0.69	1.46	0.71	1.4	0.76	1.32
	Full-form Frequency	0.57	1.76	0.44	2.29	0.71	1.42	0.43	2.34
	HFPSN	0.91	1.1	0.95	1.05	0.95	1.05	0.95	1.05
	First syllable meaning	0.99	1.01	0.93	1.07	0.99	1.01	0.94	1.07
	Phonological change	0.42	2.39	0.48	2.08	0.5	1.99	0.46	2.19
	Phonological Token Frequency (length-restraint)	0.45	2.21	0.59	1.71	0.5	2.02	0.52	1.94

Note. All frequencies were transformed to per million values and statistical analyses were conducted on log-transformed values.
LenSyll = word length in syllable; HFPSN = higher frequency phonological syllabic neighbors.