

첫 음절 토큰 빈도와 타입 빈도가 단어 및 유사 단어 어휘 판단 시간에 미치는 영향*

권 유 안†

건국대학교 언어인지 연구소

고빈도 음절로 시작되는 단어가 저빈도 음절로 시작되는 단어에 비해 어휘판단 시간이 느려지는 효과를 음절 빈도 효과라 하고, 이 효과는 시각 단어 재인에서 음절 처리를 포함한 네 수준 모형을 지지해 주는 중요한 효과로 알려져 있다. 그러나 음절 빈도는 계산 방식에 따라 타입 빈도(동일한 첫 음절을 포함한 음절 이웃 단어의 개수)와 토큰 빈도(음절 이웃 단어들의 누적 빈도)로 다르게 구분되고 각각이 단어 재인에서 다른 효과를 보일 수 있다는 주장에도 불구하고 한국어 음절 빈도 효과에서는 그 효과를 구분하지 않았다. 본 연구는 억제적 어휘 판단 시간으로 대표되는 음절 빈도 효과가 음절 타입 빈도로 설명되는지 아니면 음절 토큰 빈도로 더 잘 설명되는지를 기존 연구의 재분석을 통해 검증하였다. 분석 결과 어휘판단 시간의 증가는 음절 토큰 빈도의 증가에 의해 설명되었고, 반대로 유사 단어 자극의 어휘판단 시간의 증가는 음절 타입 빈도 및 표기 이웃 크기의 증가에 의해 설명되는 것으로 검증되었다. 이에 본 연구 결과를 바탕으로 음절 토큰 및 타입 빈도가 시각 단어 재인에서 이론적으로 어떤 의미를 갖는지를 설명하였다.

주제어 : 시각 단어 재인, 음절 빈도 효과, 음절 타입 빈도, 음절 토큰 빈도

* 본 연구는 한국연구재단 학술 연구 지원 사업 과제(35C-2011-2-H00024)의 지원으로 수행되어졌다.

† 교신저자 : 권유안, 건국대학교 언어인지 연구소, 서울시 광진구 능동로 120, yakwon75@konkuk.ac.kr

시각적으로 제시된 단어를 재인할 때 단어를 통째로 인식하는지 아니면 하위 단위로 분할하여 인식하는지에 대한 논의는 이 분야의 오랜 주제이다. 과거에는 단어를 통째로 인식한다는 주장도 있었지만(Morton, 1969), 현재의 대부분의 시각 단어 재인 모형들은 단어를 하위 단위로 분할해서 인식한다는 데 동의하고 있으며, 특히 단어의 앞부분을 단어 재인에 중요한 단위로 보고 있다(Alvarez, Carreiras, & Taft, 2001). 이러한 주장은 단어를 구성하는 하위 단위가 얼마나 자주 출현하나에 따라 단어 재인 시간이 변하는 효과를 바탕으로 한다. 특히 이 주장을 뒷받침해 주는 강력한 증거가 단어의 첫 음절의 출현 빈도를 조작한 연구들에서 지속적으로 제시되고 있다. 대표적인 연구의 사례가 스페인어 연구들로 출현 빈도가 높은 음절로 시작되는 단어가 그렇지 않은 단어에 비해 어휘 판단 시간이 증가한다는 결과이다(Carreiras, Alvarez, & Devega, 1993). 이를 음절 빈도 효과(syllable frequency effect)라 하며 이 효과는 스페인어뿐만 아니라 독일어, 프랑스어 그리고 한국어에서도 관찰되었다(Conrad & Jacobs, 2004; Kwon, Lee, Lee, & Nam, 2011; Mathey, Zagar, Doignon, & Seigneuric, 2006). Carreiras et al.(1993)의 스페인어 연구에 따르면 영어와 달리 표기 심도가 낮고 음절의 경계가 상대적으로 명확한 언어에 한해서는 자주 사용하는 고빈도 음절로 시작되는 단어가 자주 사용하지 않는 저빈도 음절로 시작되는 단어에 비해 어휘판단 시간 증가 및 오류율이 더 높다고 주장한다. 그 이유는 고빈도 음절로 시작되는 표적 단어가 저빈도 음절로 시작되는 표적 단어에 비해 자신의 첫 음절을 공유

하는 음절 이웃 단어(syllabic neighbors)를 심성 어휘집의 어휘 수준(whole word level)에서 더 많이 활성화시키게 되고, 이것은 유사한 형태의 이웃 단어들 간에 더 많은 경쟁을 일으키기 때문이라고 설명한다.

이 음절 빈도 효과의 이론적 중요성은 표기 잉여 가설(orthographic redundancy hypothesis) 또는 세 수준 상호활성화 모형이 틀렸다는 증거를 제시하는 데 있다. 이 모형은 시각 단어 재인을 위해서 시각자극의 세부특징을 처리하는 수준과 낱자를 처리하는 수준 그리고 전체 단어를 처리하는 수준을 가정한다. 그래서 두 철자 빈도만으로도 단어 안의 하위 단위를 구분할 수 있어 하위 단위를 가정할 필요가 없다는 것이다(Seidenberg & McClelland, 1989). 즉, “vodka”는 “vo”, “od”, “dk”, “ka”의 두 철자 쌍으로 이루어져 있고 각각의 두 철자 빈도(bigram frequency: 두 철자 쌍의 출현 빈도)들 중에서 “dk”의 빈도가 가장 낮아 이 부분을 음절의 경계로 인식한다는 것이다. 그러나 추후 많은 연구들이 두 철자 빈도를 동일하게 통제했음에도 불구하고 음절 빈도 효과가 관찰되었고 이에 과거 세 수준 상호활성화 모형(세부특징 수준-철자 수준-전체 단어 수준)에 음절 처리 수준을 추가한 상호활성화 모형(세부특징 수준-철자수준-음절 수준-전체 단어 수준)이 더 지지를 받고 있다.

그러나 음절 빈도 효과는 음절 빈도를 어떻게 정의하는가에 따라 다를 수 있기 때문에 보다 구체적인 연구가 요구되고 있다. 다시 말해 음절 빈도를 계산할 때 표적 단어와 동일한 첫 음절을 공유하는 단어의 개수(음절 이웃 크기)를 계산하였는지 아니면 첫 음절을

공유하는 단어의 출현 빈도의 합을 계산하였는지에 따라 다르게 계산된다. 연구자들은 전자를 음절 타입 빈도(syllabic type frequency)라 하고 후자를 음절 토큰 빈도(syllabic token frequency)라고 하며, 이렇게 음절 빈도를 나누는 이유는 음절 이웃 크기를 반영하는 음절 타입 빈도는 어휘 판단 시간을 단축시키는 효과를 보이고 반대로 음절 토큰 빈도의 증가는 어휘 판단 시간의 증가를 보이기 때문이다 (Conrad, Carreiras, & Jacobs, 2008). 이 분리된 효과를 설명하는 이론적 틀은 Grainger와 Jacobs(1996)의 MROM(multiple read-out model)에서 시작된다. 이 모형은 기본적으로는 상호 활성화 모형(interactive activation model)과 유사하지만 어휘 판단에서 “yes”반응은 두 메커니즘에 의해 결정된다고 가정한다. 첫 번째는 외측 억제 메커니즘(lateral inhibition)으로 심성 어휘집에서 활성화된 단어들 간 서로 억제적 활성화를 보내는 것을 의미하며, 외측 억제의 정도는 어휘 수준(lexical level)에서 얼마나 강력한 경쟁자가 있느냐 없느냐에 그리고 있다면 얼마나 많이 있느냐에 의해 좌우 된다. 그리고 외측 억제의 강도가 높다는 것은 강력한 경쟁자가 표적 단어의 재인을 더욱 방해하는 것을 의미하고 그 결과 어휘 판단 시간은 더 길어진다고 가정한다. 두 번째는 전체 활성화 메커니즘(global activation)으로 표적 단어가 심성어휘집 내의 표상들을 얼마나 많이 활성화시키는가와 관련있다(Grainger, Muneaux, Farioli, & Ziegler, 2005). 전체 활성화는 표적 단어의 처리가 어휘 수준에 도달하기도 전에 전체 활성화 역치 수준에 도달하면 즉각적으로 반응이 나온다고 가정한다. 다시 말해 표

적 단어가 심성 어휘집에 있는 것 같은 “yes”반응을 하게 하는 메커니즘이 전체 활성화이고 심성 어휘집의 어휘 수준에서 특정 단어를 찾아야만 “yes”반응을 하게 하는 것이 외측 억제이다.

이 모형에 따르면, 기존 억제적 음절 빈도 효과의 원인이 Carreiras et al. (1993)이 주장한 음절 이웃 크기의 증가(음절 타입 빈도의 증가)가 아니라 표적 단어 보다 더 강력한 경쟁자의 출현이라고 추정할 수 있는데, 이것은 표적 단어 보다 더 자주 출현하는 음절 이웃 단어(고빈도 음절 이웃 단어)로 음절 타입 빈도가 아닌 토큰 빈도에 더 잘 반영된다(Perea & Carreiras, 1998; Grainger, Segui, 1990). 그래서 연구자들은 고빈도 음절 이웃에 의한 단어 재인의 억제적 이 효과를 따로 음절 이웃 빈도 효과(syllable neighborhood frequency effect)라고 명명한다.

이해의 편의를 위해 한국어의 예를 들면 “목인”과 “촉감” 중 어떤 단어의 어휘 판단 시간이 더 긴지를 예측할 때, 두 단어의 음절 토큰 빈도를 계산하면 더 정확한 예측이 가능하다는 것이다. 두 단어의 출현 빈도는 동일하게 21이고 음절 타입 빈도(“목-”으로 시작되는 단어의 개수, “촉-”으로 시작되는 단어의 개수)도 20으로 동일하다. 그러나 음절 토큰 빈도는 “목인”이 79이고 “촉감”은 423으로 “촉감”이 더 크다. 두 단어의 토큰 빈도를 구성하는 이웃 단어들을 살펴보면 “목인(출현 빈도: 21)”의 음절 이웃은 “목계(출현 빈도: 10)”, “목상(출현 빈도: 8)”, “목주(출현 빈도: 6)”, “목언(출현 빈도: 5)”등으로 표적 단어 “목인”에 비해 고빈도의 음절 이웃 단어가 없다. 반대로

“촉감(출현 빈도: 21)”은 고빈도의 음절 이웃 단어 4개(‘촉각(출현 빈도: 37)’, ‘촉매(출현 빈도: 41)’, ‘촉구(출현 빈도: 127)’, ‘촉진(출현 빈도: 131)’)를 가지고 있다. 따라서 “촉감”은 “묵인”에 비해 음절 이웃들 중 “촉진”과 같은 고빈도의 이웃 단어로 부터 재인을 방해하는 외측억제 활성화를 더 많이 받을 것으로 예측이 되고 그 결과 어휘판단 시간은 “촉감”이 “묵인”에 비해 더 느려질 것으로 예측된다. 그리고 이 예측은 표기 및 음절 이웃 빈도 효과(orthographic and syllable neighborhood frequency effect)에서 모두 지지 받았다(Grainger & Segui, 1990; Kwon, Lee, Lee, & Nam, 2011; Perea & Carreiras, 1998).

음절 토큰과 타입 빈도를 조작한 국내 연구가 현재 없지만, 최근 몇몇 음절 빈도 효과와 관련된 한국어 연구들을 살펴보면, 조작된 음절 빈도는 모두 음절 타입 빈도였음에도 불구하고 일관되게 반응시간이 지연되는 억제적 효과가 나타났다. Kwon et al.(2011)의 실험 1은 단어의 출현 빈도, 첫 두 철자 빈도, 첫 음절 빈도, 표기 이웃 크기 등을 통제된 상태에서 고빈도 음절 이웃의 개수를 조작하여 고빈도 음절 이웃이 많은 단어일수록 어휘 판단 시간 및 오류율이 증가하는 것을 관찰하였다. 또한 권유안, 조혜숙, 김충명, 남기춘(2006)은 어휘 판단 과제에서 음절 빈도 효과를 관찰하였지만 역시 음절 타입 빈도만을 고려하였다. 최근 음운 음절 이웃 효과(phonological syllable neighborhood density effect)를 보고한 연구(권유안과 남기춘, 2011)에서도 음절 타입 빈도만을 조작하였고 Lee, Kwon, & Nam (2011)도 음절 빈도 효과를 보고했지만 음절 토큰 빈도는 고

려하지 않았다. 더불어 반응시간 연구가 아닌 사건관련전위(ERP: event-related potential) 연구에서도 음절 빈도 효과를 검증하기 위해 타입 빈도만을 조작하였다(Kwon, Lee, & Nam, 2011). 일반적으로 대부분의 음절 빈도 효과관련 연구들이 표적 단어로 저빈도 단어를 사용하기 때문에 음절 빈도가 높을수록 음절 이웃들은 많아지고 그 안에 고빈도의 음절 이웃이 포함될 가능성이 증가한다(Conrad et al, 2009). 때문에 위 연구들에서 억제적 어휘 판단 시간이 관찰된 이유가 음절 토큰 빈도 때문일 수 있다.

이에 본 연구는 각기 다른 목적의 연구에서 사용한 2음절의 명사 자극과 유사 단어 자극 그리고 그 어휘 판단 시간을 수집하여 음절 타입 빈도 및 토큰 빈도의 효과를 재검증하였다. 이 검증을 위해 수집된 단어 및 유사 단어(pseudo-word)들의 음절 타입 및 토큰 빈도, 그리고 두 철자 타입 및 토큰 빈도, 표기 이웃 크기, 고빈도 표기 이웃 크기들은 2007년 공개된 1500만어절의 21세기 세종계획 말뭉치(김홍규와 강범모, 2008)로부터 계산하였고 이를 독립변인 삼아 이 단어들의 어휘판단 시간과 다중회귀분석을 실시하였다.

본 분석을 통해 기대되는 결과는 단어 자극의 경우 두 철자 타입 및 토큰 빈도, 음절 타입 빈도, 표기 이웃 크기 및 고빈도 표기 이웃 크기들에 비해 음절 토큰 빈도가 어휘 판단 시간의 증가를 유의미하게 설명할 것으로 기대되며, 유사 단어의 경우 역시 음절 토큰 빈도 증가가 반응 시간 증가에 유의미한 영향을 줄 것으로 예측되었다. 단어뿐만 아니라 유사 단어도 분석을 실시한 이유는 만약 음절

토큰 빈도가 단어 재인에 영향을 준다면 유사 단어의 어휘 판단(‘단어 아님’ 반응)에도 영향을 줄 수 있기 때문이다. 일반적으로 유사 단어는 단어에 더 가까울수록 어휘 판단 시간이 느려진다. 왜냐하면 단어인줄 알고 착각했다가 다시 단어가 아니라는 추가적 판단이 필요하기 때문이다. 실제로 고빈도의 단어에서 한 철자만을 바꾸어 만든 유사 단어가 저빈도 단어에서 한 철자만을 바꾸어 만든 유사 단어에 비해 어휘판단 시간이 더 느리다(Perea, Rosa, & Gomez, 2005). 따라서 만약 음절 토큰 빈도가 유사 단어의 어휘 판단 시간의 증가를 예측하는 요인으로 판명된다면 음절 토큰 빈도가 유사 단어의 어휘성(lexicality)를 결정하는 요인으로 볼 수 있을 것이다. 이론적으로 본 연구는 음절 처리 단위를 가정하지 않는 세 수준 상호활성화 모형이 한국어 단어재인 모형에 적용가능한지 아니면 외측 억제 메커니즘을 가정하고 음절 처리 수준을 포함한 네 수준 상호활성화 모형이 적용 가능하지 타진해 볼 수 있을 것이다.

방 법

참가자 A자료는 고려대학교에서 실시한 어휘 판단 과제에 참가한 35명, B자료는 대구 가톨릭대학교에서 실시한 어휘판단 과제에 참가한 58명, C자료는 서강 대학교에서 실시한 어휘 판단 과제에 참가한 34명 모두 127명의 자료가 사용되었다.

기구 A자료는 17" CRT 모니터에 E-prime 2 프로그램(Schneider, Eschman, & Zuccolotto, 2002)

이 설치된 PC로 자극을 제시하였고 반응은 PST serial response box를 통해 수집되었다. B자료는 E-prime 1, 17" CRT 모니터를 자극 제시용으로 반응 시간은 일반 컴퓨터 키보드를 통해 얻어졌다. C자료는 자극 제시용 프로그램이 DMDX(Forster & Forster, 2002)이었고 반응은 키보드로 얻어졌다.

재료 모든 단어 자극은 2음절 명사 단어들이었다. A자료는 실험 조건이 음절 빈도와 관련 있는 것이었기 때문에 실험 자극이 아닌 메꿈 단어 64개를 수집하였다(권유안과 남기춘, 2011). B자료의 단어는 20개였고 이 자극들은 작업 기억 용량과 단어 재인 간의 관련성을 규명하기 위한 자극들이었다(이운형과 권유안, 2012). C자료는 단어 38개였고 한국어 음절체 지각에 대한 연구에서 사용된 자극이었다(Lee & Taft, 2010). 단어 수는 모두 122개였지만 겹치는 단어 8개를 제외한 114개의 단어를 수집하였다. 모든 단어와 그 어휘 판단 시간은 부록 II-1에 수록하였다. 독립변인에 대한 빈도들은 1500만 어절 21세기 세종계획 말뭉치¹⁾를 바탕으로 계산되었고 각 자료별 독립변인들에 대한 값들은 부록 I에 자세히 제시하였다.

국내외 대부분의 연구들이 지적하는 것처럼 고빈도 단어는 그 처리 과정이 매우 신속하게 이루어져 표기 이웃 크기 및 음절 빈도 관련 효과가 관찰되지 않는다(Andrews, 1992; Conrad,

1) 본 연구에서 사용한 1500만어절의 21세기 세종 계획 말뭉치는 원시 말뭉치 상태를 정제한 것이다. 현재 <http://coglab.korea.ac.kr>에서 원하는 단어 관련 빈도 값을 검색할 수 있게 공개하였다.

Carreiras, Tamm, & Jacobs, 2009, Kwon et al., 2011). 그래서 본 연구에서도 단어 출현 빈도의 평균(63.96)을 중심으로 고빈도와 저빈도 목록을 나눈(고빈도: 22개, 저빈도: 92) 후 저빈도 단어들만 분석하였다. 표 1은 실제 분석에서 사용한 저빈도 단어들의 독립변인들에 대한 속성들을 제시하였다. 표 1의 두 철자 타입 빈도는 단어의 첫 두 철자를 공유하는 단어들의 개수를 의미하고 두 철자 토큰 빈도는 그 단어들의 출현 빈도를 모두 합한 것을 의미한다(Conrad, Carreiras, & Jacobs, 2008). 예를 들어 “묘책”의 두 철자 쌍은 “ㄹㅈ”, “ㄱㅈ”, “ㅈㅈ”, “ㅈㅈ”이며 첫 두 철자 쌍 “ㄹㅈ”을 공유하는 단어들이 “묘사”, “묘지”, “묘소” 등등 약 46개가 있으며 이것을 두 철자 타입 빈도라 한다. 두 철자 토큰 빈도는 “ㄹㅈ”로 시작되는 모든 단어들의 출현 빈도들의 합을 의미하며 이 예에서 “ㄹㅈ”의 토큰 빈도는 약 805이다. 표기 이웃 크기는 표적 단어와 한 낱자만 다르고 나머지는 모두 같은 단어들의 집합을 의미한다(Coltheart, Davelaar, Jonasson, & Besner, 1977). 그리고 고빈도 표기 이웃 크기는 이 집합에서 표적 단어보다 고빈도의 표기 이웃의 크기를 의미한다(Grainger & Jacobs, 1998). 예를 들어 표적 단어가 “반사(출현빈도: 50)”인 경우 첫 음절 초성(‘ㅂ’)만 다른 표기 이웃은 “판사(출현빈도: 295)”, “간사(출현빈도: 104)”, “난사(출현빈도: 3)”, “만사(출현빈도: 51)”, “산사(출현빈도: 19)”등으로 표기 이웃의 크기는 5이다. 반면 고빈도 표기 이웃 크기는 2로 “반사”보다 고빈도의 표기 이웃 단어인 “판사”, “간사”가 그것이다.

모든 유사 단어의 첫 음절은 자음-모음-자

음으로 이루어져 있으며 가급적이면 한국어에서 사용되고 있는 음절로 시작되는 자극만 선정하였다. 예를 들어 “좁옥”은 유사 단어 범주이지만, “쫙쫙”과 같은 비단어(illegal word)는 제외했다. 유사 단어의 출현 빈도는 0이고 따라서 단어 출현 빈도 요인과 고빈도 표기 이웃은 존재하지 않는다. 그래서 분석에 사용한 독립 변인은 표기 이웃 개수, 두 철자 타입 및 토큰 빈도, 음절 타입 및 토큰 빈도였다. 유사 단어 자극의 자세한 자료별 독립 변인의 값들은 부록 I에 제시하였고 유사 단어와 어휘 판단 시간은 부록 II-2에 수록하였다. 표 1의 유사 단어의 값들은 수집된 자료를 통합한 것이다.

절차 모든 자료는 자극을 제시하고 그 자극이 단어인지 아닌지를 판단시키는 어휘 판단 과제에서 얻어졌다. 자극의 제시 순서는 처음 응시점을 제시하고 즉각적 또는 약간의 간격을 두고 표적 단어를 제시하고 이때 실험참가자의 반응을 받고, 다음 수행으로 넘어가는 것이었다. A와 B자료의 어휘 판단 시간은 자극이 제시되고 반응이 이루어질 때까지 자극이 제시되는 어휘 판단 과제에서 얻어진 것이고 C자료는 자극이 약 300ms동안만 제시되고 참가자가 이를 어휘 판단하는 과제였다. 각 연구마다 자극 제시 시간 및 간격 그리고 지시 사항 등에는 약간의 차이가 있었지만 자극 제시 후 실험 참가자의 반응(단어이면 “yes” 단어가 아니면 “no”)이 이루어질 때까지의 시간을 반응시간으로 기록한 점 그리고 자극이 제시되면 단어인지 아닌지를 빠르고 정확하게 판단하라는 지시문도 일치하였다.

분석방법 분석에서 두 철자 및 음절 타입 빈도와 토큰 빈도를 한 회귀방정식의 독립 변인으로 사용하지 않았는데, 그 이유는 두 철자 토큰 빈도와 타입 빈도의 상관($r=.92, p<.0001$)과 음절 토큰과 타입 빈도의 상관($r=.64, p<.0001$)이 너무 높아 다중공선성의 문제가 있기 때문이었다. 그리고 각 회귀 모형에서 사용된 독립변인들의 TOL 및 VIF 값은 모두 기준 값에서 멀리 떨어져 있어 다중공선성의 문제는 없었다(자세한 상관표 및 다중공선성 지표 TOL 및 VIF에 대한 표는 부록 III-1과 2에 제시하였다). 물론 두 철자 타입 및 토큰은 주요 관심사가 아니기 때문에 두 변인들의 평균값을 하나의 요인으로 묶어서 분석할 수도 있었겠지만, 두 철자 토큰과 타입 빈도를 각각

분리하는 것이 본 연구의 의도에 더 적절하다고 판단하여 둘을 나누었다. 종속치는 정답을 낸 반응시간만을 모아서 사용하였다. 세 연구의 단어 목록의 정확률은 평균 85%였다. 그리고 유사 단어의 정답은 단어가 아니라고 정확하게 판단한 반응들만 모았고 평균 정확률은 약 78%였다. 정제된 반응시간과 추출된 독립 변인들 간에 독립적인 4회의 다중회귀분석을 실시하였다.

단어 자극 분석에서 첫 번째 다중회귀분석(모형1-1)에 사용된 독립 변인은 표기 이웃 크기, 고빈도 표기 이웃 크기, 두 철자 타입 빈도, 음절 타입 빈도였다. 두 번째 다중회귀분석(모형1-2)에서는 표기 이웃 크기, 고빈도 표기 이웃 크기, 두 철자 토큰 빈도, 음절 토큰

표 1. 단어 및 유사 단어 자극의 독립변인 속성표

자극 유형	독립 변인	평균	표준편차
단어자극 (저빈도: N=92)	단어 출현 빈도	17	14.31
	표기 이웃 크기	3	2.59
	고빈도 표기 이웃 크기	2	2.77
	두 철자 타입 빈도	387	217.93
	두 철자 토큰 빈도	23250	19845.52
	음절 타입 빈도	79	51.19
	음절 토큰 빈도	4155	4504.80
유사 단어 자극 (N=140)	표기 이웃 크기	1	1.67
	두 철자 타입 빈도	359	219.00
	두 철자 토큰 빈도	24090	19196.28
	음절 타입 빈도	38	40.37
	음절 토큰 빈도	1852	3110.57

주. 모든 빈도는 1500만 어절의 21세기 세종계획에서 발표한 말뭉치에서 계산된 것이다. 계산의 기준이 되는 단어의 부분은 단어 빈도를 제외하고 모두 단어의 첫 음절이었다. 토큰 빈도를 log로 치환하여 제시하는 것이 일반적이지만 누적빈도라는 것을 명시하기 위해 log로 치환하지 않았다.

빈도를 독립변인으로 사용하였다. 세 번째 다중회귀분석(모형1-3)에서는 표기 이웃 크기, 고빈도 표기 이웃 크기, 두 철자 타입 빈도, 음절 토큰 빈도를 독립변인으로 하였다. 네 번째(모형1-4)는 표기 이웃 크기, 고빈도 표기 이웃 크기, 두 철자 토큰 빈도, 음절 타입 빈도를 독립변인으로 하였다. 유사 단어 반응 시간 분석에서 첫 번째 다중회귀분석(모형2-1)에 사용된 요인은 표기 이웃 크기, 두 철자 타입 빈도, 음절 타입 빈도였다. 두 번째(모형2-2)는 표기 이웃 크기, 두 철자 토큰 빈도, 음절 토큰 빈도였고, 세 번째(모형2-3)는 표기 이웃 크기, 두 철자 타입 빈도, 음절 토큰 빈도 였다. 마지막 분석(모형2-4)의 독립변인은 표기 이웃

크기, 두 철자 토큰 빈도, 음절 타입 빈도였다.

결 과

단어 자극 분석 결과 네 개의 다중회귀분석 모형에 대한 검증에서 통계적으로 유의미한 모형은 없었다. 그러나 고빈도 표기 이웃 크기, 표기 이웃 크기, 두 철자 타입 빈도, 음절 토큰 빈도를 독립변인으로 한 모형1-2가 유의수준에 근접하였다($R^2=.087$, *adjusted* $R^2=.045$, $F(4, 87) = 2.09$, $p = .08$). 그러나 모형1-2에 대한 단계적 회귀분석에서 음절 토큰 빈도가 어휘 판단 시간의 증가를 유의미하게 설명하였다($R^2=.05$, $F(1, 90)=5.46$, $p<.02$). 또한 모형1-4

표 2. 어휘 판단 시간과 독립 변인 간의 다중회귀 분석 결과
(변인의 괄호 안 숫자는 단계적 회귀 분석에서 선택된 순서를 의미함)

모형	변인	β 계수	SE	표준 β 계수	t 값	p
모형1-1	표기 이웃 크기	.594	5.828	.014	-.33	.74
	고빈도 표기 이웃 크기	-1.802	5.415	-.046	.10	.91
	음절 타입 빈도	-.011	.267	-.005	-.04	.96
	두 철자 타입 빈도	-.020	.061	-.042	-.34	.73
모형1-2	표기 이웃 크기	.201	5.634	.004	.04	.74
	고빈도 표기 이웃 크기	-1.672	5.221	-.043	-.32	.97
	음절 토큰 빈도(1)	.007	.002	.300	2.61	.01**
	두 철자 토큰 빈도	-.000	.000	-.137	-1.18	.23
모형1-3	표기 이웃 크기	.488	5.863	.011	.08	.93
	고빈도 표기 이웃 크기	-1.717	5.449	-.044	-.03	.75
	음절 타입 빈도	-.067	.259	.009	-.26	.79
	두 철자 토큰 빈도	.000	.000	-.032	.08	.93
모형1-4	표기 이웃 크기	-.115	5.554	-.002	-.02	.98
	고빈도 표기 이웃 크기	-1.021	5.179	-.026	-.20	.84
	음절 토큰 빈도(1)	.007	0.056	.323	2.82	.005**
	두 철자 타입 빈도(2)	-.092	0.002	-.188	-.64	.10

주. ** $p<.01$

에 대한 단계적 회귀분석(고빈도 표기 이웃 크기, 표기 이웃 크기, 두 철자 토큰 빈도, 음절 토큰 빈도)에서도 음절 토큰 빈도만이 유의미하였다($R^2=.05$, $F(1, 90)=5.46$, $p<.02$). 각 모형의 자세한 분석결과는 표 2에 제시하였다. 모든 결과는 독립변인의 값들을 log로 치환해도 동일하였다.

분석 결과 저빈도 단어들의 어휘판단 시간에 가장 큰 영향을 주는 하위 단위 요인은 음절 토큰 빈도였다. 특히 어휘 판단 시간은 음절 토큰 빈도의 증가에 따라 증가하는 양상이 나타났기에 억제적 음절 빈도 효과가 음절 타입 빈도가 아닌 토큰 빈도에서 기인한다는 연구와 일치한다(Conrad et al., 2008; Conrad et al., 2009). 그러나 음절의 토큰 빈도와 타입 빈도 간의 상관 계수가 높은 것($r=.63$, $p<.0001$)을 고려하면 음절 타입 빈도만을 조작한 한국어 연구들에서 관찰된 억제적 음절 빈도 효과에는 음절 토큰 빈도가 내재되어 있지 않나 추측된다.

유사 단어 자극 분석 결과 모형2-1의 독립변인은 표기 이웃 크기, 두 철자 타입 빈도, 음절 타입 빈도였다. 이 분석 모형은 유의미하였다($R^2=.18$, $adjusted R^2=.16$, $F(3, 136)=10.46$, $p<.0001$). 모형2-1의 단계적 회귀분석에서 가장 먼저 선택된 요인은 음절 타입 빈도였고($R^2=.10$, $F(1, 138)=15.91$, $p<.0001$), 두 번째는 표기 이웃 크기였고($R^2=.05$, $F(2,137)=9.48$, $p<.002$), 세 번째는 두 철자 타입 빈도($R^2=.02$, $F(3,136)=4.36$, $p<.03$)이었다. 모형2-2에서는 표기 이웃 크기, 두 철자 토큰 빈도, 음절 토큰 빈도가 독립변인이었다. 이 모형도 유의미하

였다($R^2=.08$, $adjusted R^2=.06$, $F(3, 136)=3.96$, $p<.009$). 단계적 회귀분석에서 첫 번째로 채택된 요인은 표기 이웃 크기였고($R^2=.04$, $F(1, 138)=7.21$, $p<.008$), 두 번째로 선택된 요인은 음절 토큰 빈도였고($R^2=.03$, $F(2,137)=4.49$, $p<.03$), 두 철자 토큰 빈도는 반응시간에 영향을 끼치지 못했다. 모형2-3에서는 독립변인이 표기 이웃 크기, 두 철자 타입 빈도, 음절 토큰 빈도였다. 이 모형은 유의미하였으며($R^2=.08$, $adjusted R^2=.06$, $F(3, 136)=4.09$, $p<.008$), 단계적 회귀분석에서 선택된 독립변인은 표기 이웃 크기($R^2=.04$, $F(1,138)=7.21$, $p<.008$)와 음절 토큰 빈도였다($R^2=.03$, $F(2,137)=5.94$, $p<.003$). 모형2-4의 독립변인은 표기 이웃 크기, 두 철자 토큰 빈도, 음절 토큰 빈도였다. 이 모형도 역시 유의미하였으며($R^2=.17$, $adjusted R^2=.15$, $F(3, 136)=9.61$, $p<.0001$), 단계적 회귀분석에서 음절 타입 빈도가 가장 먼저 선택되고($R^2=.10$, $F(1,138)=15.91$, $p<.0001$), 다음으로 표기 이웃 크기가 선택되었다($R^2=.05$, $F(2,137)=9.48$, $p<.002$). 역시 모형2-4에서도 두 철자 토큰 빈도는 반응 시간에 아무런 영향을 주지 못하는 것으로 판명되었다. 단어 자극 분석과 마찬가지로 log로 치환된 값을 사용한 분석에서도 결과는 동일하였다. 유사 단어 분석에서 어휘판단 시간에 주요한 영향을 준 요인은 주로 표기 이웃 크기와 음절 타입 빈도였다. 예측과는 다르게 표기 이웃 크기가 음절 토큰 빈도 보다 더 유의미하게 어휘 판단 시간에 영향을 끼치는 것으로 나타났다. 음절 토큰 빈도도 유의미한 영향을 끼치는 것으로 나타났으나 그 영향이 표기 이웃 크기와 음절 타입 빈도에 비하면 작았다.

표 3. 어휘판단 시간과 독립 변인간의 다중회귀 분석 결과

(변인의 괄호 안 숫자는 단계적 회귀 분석에서 선택된 순서를 의미함)

모형	변인	β 계수	SE	표준 β 계수	t 값	P
모형2-1	표기 이웃 크기(2)	13.132	4.102	.248	3.20	.001***
	음절 타입 빈도(1)	1.043	.217	.397	4.79	.0001***
	두 철자 타입 빈도(3)	-0.083	.039	-.173	-2.09	.03*
모형2-2	표기 이웃 크기(1)	11.912	4.354	.224	2.74	.006***
	음절 토큰 빈도(2)	.004	.002	.178	2.12	.03*
	두 철자 토큰 빈도	-0.000	.000	-.024	-.03	.76
모형2-3	표기 이웃 크기(1)	11.912	4.350	.226	2.74	.006***
	음절 토큰 빈도(2)	.004	.002	.181	2.12	.03*
	두 철자 타입 빈도	-.025	.039	-.054	-.65	.51
모형2-4	표기 이웃 크기(2)	12.833	4.131	.242	3.11	.002**
	음절 타입 빈도(1)	.992	.218	.378	4.54	.0001***
	두 철자 토큰 빈도	-0.000	.000	-.124	-1.49	.13

주. * $p < .05$, ** $p < .01$, *** $p < .001$

표 4. 8개 다중회귀분석 결과 분석 요약표

자극 종류	유의미한 독립 변인	어휘 판단 시간의 경향
단어 자극 분석	음절 토큰 빈도	'yes'반응 시간의 증가
유사 단어 자극 분석	음절 타입 빈도 표기 이웃 크기	'no'반응 시간의 증가

두 분석 결과를 요약하면 표 4와 같다. 표 4는 총 8개의 다중회귀분석에서 유의수준 .01 이하로 2회 이상 어휘판단 시간의 증가를 유의미하게 설명한다고 선택된 독립변인만을 간추려 제시하였다.

논 의

본 연구의 목적은 음절 빈도 효과의 원인이 음절 토큰 빈도 인지 아니면 타입 빈도 인지를 검증하여 단어 재인 모형 중 세 수준 상호

활성화 모형이 맞는지 아니면 음절 처리를 포함한 네 수준 상호활성화 모형이 맞는지를 검증하는 것이었다. 이를 위해 각기 다른 연구 목적의 단어 자극 및 유사 단어와 그 반응시간을 수집하여 저빈도 단어들과 유사 단어 자극들의 어휘 판단 시간과 그 자극들의 음절 타입 빈도와 토큰 빈도를 포함한 단어 하위 요인들 간의 다중회귀분석을 실시하였다. 단어 자극 분석 결과 예측대로 음절 토큰 빈도가 증가할수록 어휘 판단 시간이 증가하는 추세가 나타났다. 하지만 유사 단어 자극 분석

에서는 음절 타이핑 빈도와 더불어 표기 이웃 크기가 유사 단어의 어휘 판단 시간에 영향을 끼친 것으로 나타났다. 따라서 음절 처리를 가정하지 않는 세 수준 상호활성화 모형은 음절 토큰 빈도 효과를 설명할 수 없으므로 본 연구의 결과는 음절 처리를 포함한 네 수준 상호활성화 모형을 지지하는 것이다.

본 연구의 결과는 음절 토큰 빈도가 스페인어의 음절 빈도 효과를 설명해 주는 주요 요인이라는 Conrad, Carreriras, & Jacobs (2008)의 결과와도 일치한다. 이들에 따르면 토큰 빈도 증가에 따른 어휘 판단 시간의 증가는 단어가 표상된 어휘 수준에서 외측 억제 메커니즘이 반영된 결과이고, 반대로 타이핑 빈도 증가에 따른 어휘 판단 시간의 축진은 어휘 수준보다는 그 이전 단계의 처리에 영향을 주는 요인이기 때문에 외측 억제가 아닌 전체 활성화를 반영한 결과라고 설명하였다.

위 설명을 본 연구 결과에 적용하면 단어 자극의 경우 단어이기 때문에 심성 어휘집의 어휘 수준까지 처리가 될 것이다. 만약 표적 단어가 음절 토큰 빈도가 높다면 그 표적 단어 보다 고빈도의 음절 이웃 단어가 더 많이 있을 수 있고 따라서 강한 경쟁자가 어휘 수준에서 활성화 될 것이다. 그 결과 그 표적 단어는 경쟁자로부터 강한 외측 억제를 받고 반응 시간이 증가된 것으로 해석된다. 반대로 유사 단어는 단어가 아니기 때문에 심성 어휘집의 어휘 수준에 도달하지 못한다. 그래서 외측 억제 활성화와는 무관하다. 대신 한 유사 단어를 구성하는 음절이나 철자가 전체 활성화 수준을 결정할 것이다. 본 연구 결과에 의하면 음절 타이핑 빈도와 한 철자를 제외한

나머지 철자가 같은 표기 이웃 크기(이것 역시 타이핑 빈도임)의 증가가 “no”반응을 지연시키는 데 주요한 영향을 끼친 것으로 나타났다. 이 결과는 표적 유사 단어가 음절 타이핑 빈도가 높거나 또는 표기 이웃이 커서 “yes”반응을 촉발시킬 수 있는 역치 수준에 도달하지만 단어가 아니므로 이 역치 수준을 다시 설정하고 “no”반응으로 전환해야하는 과정이 추가되어 반응 시간이 지연된 것으로 해석된다(Carreiras et al., 1997; Holcomb, Grainger, & O’Rourke, 2002). 만약 이 과정이 아니라 유사 단어가 어휘 수준의 단어를 활성화 시켰다면 단어 자극 분석과 동일하게 음절 토큰 빈도가 본 연구에서 유의미한 독립변인으로 도출되었을 것이다. 따라서 유사 단어의 어휘성을 결정한 요인은 음절 토큰 빈도가 아니라 음절 타이핑 빈도와 표기 이웃 크기인 것으로 밝혀졌다.

본 연구의 결과는 음절 빈도 효과가 음절 토큰 빈도에 의해 설명된다는 기존의 연구들에 또 하나의 증거를 실험 연구가 아닌 통계적 결과로 제시하고 있다. 이 통계적 지지 증거는 시각 단어재인 분야에서 다년간 진행되고 있는 이론적 쟁점인 표기 잉여 가설(orthographic redundancy hypothesis) 또는 세 수준 상호활성화 모형이 옳은지 아니면 음절 처리를 포함한 모형(interactive activation model incorporating syllabic parser) 또는 네 수준 모형이 옳은지에 대한 논쟁에서 후자를 지지하는 결과이며 또한 그 주장을 일반화시킬 수 있는 근거이다. 그러나 본 연구에서 사용된 자극의 수가 다른 국외 연구들에 비해 적은 점(Balota et al.(2004)은 무려 2,428개의 자극을 사용했음), 명사 자극만을 분석한 점, 수집된 자

극이 모두 어휘 판단 과제였다는 점은 추가적 연구를 요구하게 한다.

참고문헌

- 권유안, 조혜숙, 김충명, 남기춘 (2006). 한국어 시각단어재인에서 나타나는 이웃효과. *말소리*, 60, 30-45.
- 권유안, 남기춘 (2011). 한글 음절 이웃 효과에서 한자어 형태소의 영향: 표기 및 음운 이웃과 한자어 이웃과의 관련성 중심으로. *한국심리학회지: 인지 및 생물*, 23(3), 301-319.
- 김흥규, 강범모. (2008). 한국어 단어와 형태소의 사용 빈도 1500만 어절의 세종형태 의미 분석 말뭉치 기반. 서울: 고려대학교 민족문화연구원.
- 이윤형, 권유안 (2012). 작업기억의 개인차와 문장의 이해. *Journal of the Korean Data Analysis Society*, 14(2B), 825-836.
- Alvarez, C. J., Carreiras, M., & Taft, M. (2001). Syllables and morphemes: Contrasting frequency effects in Spanish. *Journal of Experimental Psychology-Learning Memory and Cognition*, 27(2), 545-555.
- Andrews, S. (1992). Frequency and Neighborhood Effects on Lexical Access - Lexical Similarity or Orthographic Redundancy. *Journal of Experimental Psychology-Learning Memory and Cognition*, 18(2), 234-254.
- 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), 283-316.
- Carreiras, M., Alvarez, C. J., & Devega, M. (1993). Syllable Frequency and Visual Word Recognition in Spanish. *Journal of Memory and Language*, 32(6), 766-780.
- Coltheart, M., Davelaar, E., Jonasson, J. T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and Performance, VI* (pp.535-555). Hillsdale, NJ: Erlbaum.
- 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), 461-479.
- Conrad, M., & Jacobs, A. M. (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.
- 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.
- Forster, K. I., & Forster, J. C. (2002). DMDX display software.

- Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review*, 103(3), 518-565.
- Grainger, J., Muneaux, M., Farioli, F., & Ziegler, J. C. (2005). Effects of phonological and orthographic neighbourhood density interact in visual word recognition. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology*, 58(6), 981-998.
- Grainger, J., & Segui, J. (1990). Neighborhood Frequency-Effects in Visual Word Recognition - a Comparison of Lexical Decision and Masked Identification Latencies. *Perception & Psychophysics*, 47(2), 191-198.
- Kim, J., Kwon, Y., & WWW, K. (2011). The Syllable Frequency Effect in Semantic Categorization Tasks in Korean. *Ksii Transactions on Internet and Information Systems*, 5(10), 1879-1890.
- Kwon, Y., Lee, Y., & Nam, K. (2011). The different P200 effects of phonological and orthographic syllable frequency in visual word recognition in Korean. *Neuroscience Letters*, 501(2), 117-121.
- Kwon, Y., Lee, Y., Lee, K., & Nam, K. (2011). The inhibitory effect of phonological syllables, rather than orthographic syllables, as evidenced in Korean lexical decision tasks. *Psychologia*, 54, 1-14.
- Mathey, S., Zagar, D., Doignon, N., & Seigneuric, A. (2006). The nature of the syllabic neighborhood effect in French, *Acta Psychologica*, 12, 372-393.
- Morton, J. (1969). Interaction of Information in Word Recognition. *Psychological Review*, 76(2), 165-178.
- Perea, M., & Carreiras, M. (1998). Effects of syllable frequency and syllable neighborhood frequency in visual word recognition. *Journal of Experimental Psychology-Human Perception and Performance*, 24(1), 134-144.
- Perea, M., Rosa, E., & Gomez, C. (2005). The frequency effect for pseudowords in the lexical decision task. *Perception & Psychophysics*, 67(2), 301-314.
- Rice, G. A., & Robinson, D. O. (1975). Role of Bigram Frequency in Perception of Words and Nonwords. *Memory & Cognition*, 3(5), 513-518.
- Seidenberg, M. S. (1989). Reading complex words. In G. Carlson & M. Tannenhaus (Eds.), *Linguistic structure in language processing*. Boston: Kluwer Academic.
- Schneider, W., Eschman, A., & Zuccolotto, A. (2002). E-prime Reference Guide. *Pittsburgh: Psychology Software Tools, Inc.*

1 차원고접수 : 2012. 8. 16
수정원고접수 : 2012. 9. 16
최종게재결정 : 2012. 10. 5

The Dissociation of Syllabic Token and Type Frequency Effect in Lexical Decision Task

Youan Kwon

Language & Cognition Research Center of Konkuk University

Syllable frequency effect plays an important role in finding evidence supporting the hypothesis, incorporating syllable unit between letter and whole word level. The syllable frequency effect is defined as the inhibitory effect of words starting with high syllable frequency in lexical decision tasks. According to some studies, syllable frequency can be divided into two values. First is the type frequency which indicates the number of syllabic neighbors sharing the first syllable of a target word. Second is the token frequency which means the accumulated word frequency of the syllabic neighbors. Although type and token syllable frequency can make different effect, there is no evidence in Korean. This study conducted the meta-analyses to access whether the inhibitory syllable frequency effect could be expected by syllabic type or token frequency. The results showed that the syllabic token frequency played a critical role in the inhibitory effect in Korean syllable frequency effect and we explained the role of token and type frequency in visual word recognition model.

Key words : visual word recognition, syllabic token frequency, syllabic type frequency, syllable frequency effect

부록 I

각 자료별 독립 변인들의 평균 빈도 및 표준 편차

	A 자료	B 자료	C 자료
단어 자극			
단어 출현 빈도	17(11.66)	196(287.42)	75(167.55)
표기 이웃 크기	4(2.66)	2(1.23)	3(2.68)
고빈도 표기 이웃 크기	3(3.03)	0(0.58)	1(1.86)
두 철자 타입 빈도	391(244.95)	368(240.80)	414(153.46)
두 철자 토큰 빈도	23362(22848.64)	24228(21458.86)	26246(13498.43)
음절 타입 빈도	79(54.06)	72(59.87)	86(45.01)
음절 토큰 빈도	3326(3212.73)	2948(3340.35)	6743(5828.11)
유사 단어 자극			
표기 이웃 크기	2(2.01)	0.7(1.20)	1(1.81)
두 철자 타입 빈도	298(198.24)	338(292.24)	441(166.52)
두 철자 토큰 빈도	17491(16787.71)	25262(27056.70)	29517(13744.44)
음절 타입 빈도	85(87.19)	80(96.45)	64(76.21)
음절 토큰 빈도	1617(3826.75)	1659(2286.72)	2280(3218.23)

주. 모든 빈도는 1500만 어절의 21세기 세종계획에서 발표한 말뭉치에서 계산된 것이다. 계산의 기준이 되는 단어의 부분은 단어 빈도를 제외하고 모두 단어의 첫 음절이었다. 토큰 빈도를 log로 치환하여 제시하는 것이 일반적이지만 누적빈도라는 것을 명시하기 위해 log로 치환하지 않았다. 두 철자 타입 빈도는 단어의 첫 두 표기가 연달아 나온 단어들의 개수이고 두 철자 토큰 빈도는 동일 두 표기를 가진 단어들의 누적 빈도이다. 이 두 용어 외의 설명은 권유안과 남기춘(2011)에 상세히 실려 있다.

부록 II-1

자극 단어 목록 및 평균 어휘 판단 시간

A 자료 자극 단어 (평균 어휘 판단 시간)			
감색 (619.7)	백미 (563.7)	약경 (615.0)	폭군 (578.8)
광원 (705.9)	백호 (611.1)	양민 (595.5)	학도 (669.4)
광학 (614.7)	변기 (568.3)	역도 (600.6)	혈기 (632.9)
권역 (688.4)	병정 (594.2)	용맹 (556.1)	훈계 (632.0)
근력 (588.8)	복통 (597.2)	윤택 (581.5)	족속 (680.9)
단오 (578.8)	분과 (665.5)	응수 (657.5)	탐방 (612.2)
등잔 (616.7)	산보 (627.7)	짓갈 (646.2)	통소 (703.5)
뗏목 (681.1)	산삼 (588.6)	종가 (626.5)	한식 (552.8)
막후 (676.4)	상감 (725.1)	중절 (668.8)	낙선 (617.1)
명물 (577.7)	석권 (586.7)	참관 (598.5)	반원 (650.1)
뉘음 (637.3)	설달 (680.5)	천도 (676.6)	옷장 (610.8)
민가 (632.6)	식견 (590.9)	청부 (672.3)	용왕 (576.5)
반경 (642.6)	실추 (660.9)	축사 (610.1)	탐닉 (614.8)
발매 (739.3)	심술 (581.1)	축출 (638.6)	뒤틀 (651.6)
발압 (601.6)	약덕 (585.7)	특강 (620.7)	
백마 (565.7)	안치 (680.2)	팽이 (606.6)	

B 자료 자극 단어(평균 어휘 판단 시간)			
압력 (622.1)	폭로 (620.0)	막내 (615.8)	침략 (693.5)
신뢰 (621.8)	신랑 (611.5)	책임 (689.2)	녹용 (694.5)
약물 (683.0)	잡음 (773.2)	특허 (735.5)	숙녀 (572.0)
명령 (652.3)	입학 (627.2)	혈연 (727.6)	후연 (933.0)
난로 (701.0)	중력 (623.4)		

C 자료 자극 단어(평균 어휘 판단 시간)			
논쟁 (630.4)	반란 (666.6)	작물 (738.7)	
난리 (643.5)	작동 (668.7)	백반 (840.1)	
신비 (646.1)	환갑 (731.0)	극락 (862.8)	
국밥 (639.8)	논란 (760.3)	국난 (870.2)	
논의 (676.3)	국모 (783.7)	독려 (830.6)	
반복 (686.0)	난타 (776.5)	국채 (887.7)	
폭소 (689.7)	만사 (771.9)	국관 (953.3)	
독기 (723.6)	폭리 (805.0)	만루 (971.4)	
논리 (713.7)	국책 (858.1)	녹물 (1095.)	
국물 (720.3)	녹봉 (875.7)	환락 (849.2)	
신라 (722.5)	국선 (755.1)	환관 (826.1)	
녹말 (688.4)			

부록 II-2

자극 유사 단어 목록 및 평균 어휘 판단 시간

A 자료 유사 단어 자극(평균 어휘 판단 시간)			
감절 (775.23)	단포 (814.77)	손계 (724.45)	탕류 (740)
격수 (679.36)	당박 (748.31)	실뢰 (834.53)	탕성 (726)
검람 (699.65)	택결 (706.68)	실람 (820.36)	판관 (829.47)
곳청 (640.16)	덩색 (817.76)	안명 (902.81)	핀지 (745.16)
몸섭 (733.7)	둔공 (757.59)	압고 (790.61)	함위 (798.88)
관송 (802.5)	둘색 (808.14)	압프 (674.00)	향각 (814.39)
웅골 (769.21)	명유 (733.58)	엔행 (664.45)	행곡 (743.06)
김양 (735.87)	맨수 (761.67)	웃넘 (727.28)	흙비 (730.75)
닝적 (709.65)	면롱 (681.18)	음시 (709.87)	혼랑 (810.95)
남랑 (782.31)	면막 (827.07)	음마 (657.81)	혼렬 (785.06)
남장 (843.21)	물결 (859.31)	인력 (833.96)	홍자 (872.95)
닝박 (729.73)	땀금 (650.8)	잔은 (738.15)	
줍자 (763.64)	병면 (832.78)	쟁치 (794.16)	
님관 (777.70)	민샤 (659.38)	촌세 (814.87)	

B 자료 유사 단어 자극(평균 어휘 판단 시간)			
정작 (843.80)	군실 (1275.3)	탱잔 (860.70)	산릴 (832.39)
싱류 (857.82)	격문 (942.76)	형자 (715.15)	암닐 (770.12)
락셀 (796.66)	만샤 (784.50)	흑원 (1025.6)	
행규 (926.05)	긴태 (866.88)	냉석 (1116.6)	

C 자료 유사 단어 자극(평균 어휘 판단 시간)			
준망 (809.59)	굴면 (789.53)	줍얼 (734.57)	춧임 (726.42)
곽춤 (652.12)	준방 (741.20)	문결 (757.13)	돌길 (750.60)
묵긴 (700.88)	곽름 (618.04)	낙분 (917.81)	남목 (684.95)
낙견 (767.29)	묵빈 (690.86)	죽문 (702.51)	춤옥 (626.49)
송집 (732.07)	낙견 (775.83)	각련 (789.65)	갈역 (716.03)
밤늑 (733.59)	송넘 (630.23)	싱결 (750.52)	식녕 (661.61)
몽족 (747.07)	몽잔 (636.51)	논땀 (755.56)	종립 (658.17)
랑격 (613.84)	잠폭 (707.35)	줍역 (711.30)	묵린 (657.15)
긴성 (656.63)	군막 (725.37)	문겹 (715.45)	남늑 (644.15)
몽닌 (637.58)	잔녁 (807.54)	몽갈 (605.40)	춤곡 (641.79)
울집 (679.85)	형산 (948.41)	갑푼 (620.29)	갈늑 (810.69)
굴면 (658.78)	춤싱 (673.37)	군땀 (642.74)	식명 (930.24)
밤룩 (718.82)	둑릭 (702.77)	잔녕 (661.54)	종겹 (704.05)
몽늑 (706.15)	낙뵘 (685.27)	형삭 (726.53)	묵닌 (675.52)
랑떡 (716.52)	죽몽 (703.22)	춤신 (596.49)	잔뵘 (678.31)
긴녕 (653.25)	각령 (739.76)	둑링 (677.64)	잔련 (726.68)
몽딘 (624.44)	싱견 (658.14)	잔굽 (749.03)	헛망 (803.21)
울땀 (669.82)	논밍 (625.17)	잔견 (781.83)	춧땀 (581.18)
			돌넙 (662.29)

부록 III-1

단어 자극 독립변인들 간의 상관행렬표

	고빈도 표기 이웃 개수	표기 이웃 개수	두 철자 타입 빈도	두 철자 토큰 빈도	음절 타입 빈도	음절 토큰 빈도
고빈도 표기 이웃 개수		.64***	.09	.04	.19	.01
표기 이웃 개수			.13	.14	.22	.06
두 철자 타입 빈도				.92***	.42**	.44**
두 철자 토큰 빈도					.46**	.44**
음절 타입 빈도						.64***
음절 토큰 빈도						

주. ** < .01, *** < .001

단어 자극 분석 모형에서의 다중공선성
지표 TOL(tolerance) 및 VIF(variance inflation factor)값

모형	독립변인	TOL	VIF
모형1-1	고빈도 표기 이웃 개수	.58	1.70
	표기 이웃 개수	.57	1.73
	두 철자 타입 빈도	.72	1.37
	음절 타입 빈도	.70	1.42
모형1-2	고빈도 표기 이웃 개수	.58	1.71
	표기 이웃 개수	.57	1.74
	두 철자 토큰 빈도	.78	1.27
	음절 토큰 빈도	.80	1.24
모형1-3	고빈도 표기 이웃 개수	.58	1.70
	표기 이웃 개수	.58	1.72
	두 철자 타입 빈도	.79	1.26
	음절 토큰 빈도	.79	1.25
모형1-4	고빈도 표기 이웃 개수	.57	1.72
	표기 이웃 개수	.57	1.75
	두 철자 토큰 빈도	.77	1.29
	음절 타입 빈도	.74	1.33

주. 다중공선성을 평가하는 TOL값은 .1미만이면 VIF값은 10을 초과하면 해당 변수의 다중공선성의 문제가 있는 것으로 간주한다. 각 모형에서 다중공선성에 문제가 있는 변인은 없는 것으로 검증되었다.

부록 III-2

단어 자극 독립변인들 간의 상관행렬표

	표기 이웃 개수	두 철자 타입 빈도	두 철자 토큰 빈도	음절 타입 빈도	음절 토큰 빈도
표기 이웃 개수		.01	-.01	-.05	-.01
두 철자 타입 빈도			.92***	.36**	.13
두 철자 토큰 빈도				.35**	.18
음절 타입 빈도					.73***
음절 토큰 빈도					

주. ** < .01, *** < .001

유사 단어 자극 분석 모형에서의 다중공선성
지표 TOL(tolerance) 및 VIF(variance inflation factor)값

모형	독립변인	TOL	VIF
모형2-1	표기 이웃 개수	.99	1.00
	두 철자 타입 빈도	.86	1.15
	음절 타입 빈도	.86	1.15
모형2-2	표기 이웃 개수	.99	1.00
	두 철자 토큰 빈도	.96	1.03
	음절 토큰 빈도	.96	1.03
모형2-3	표기 이웃 개수	.99	1.00
	두 철자 타입 빈도	.98	1.02
	음절 토큰 빈도	.98	1.02
모형2-4	표기 이웃 개수	.99	1.00
	두 철자 토큰 빈도	.87	1.14
	음절 타입 빈도	.87	1.14

주. 다중공선성을 평가하는 TOL값은 .1미만이면 VIF값은 10을 초과하면 해당 변수의 다중공선성의 문제가 있는 것으로 간주한다. 각 모형에서 다중공선성에 문제가 있는 변인은 없는 것으로 검증되었다.