

명명 과제에서 음절 토큰 및 타입 빈도 효과*

권 유 안†

건국대학교 다문화 다언어 연구소

음절 빈도 효과란 고빈도 음절로 시작되는 단어가 저빈도 음절로 시작되는 단어에 비해 어휘 판단 속도가 느리며 어휘 판단 오류율도 증가하는 효과를 의미한다. 이 효과를 유발하는 원인은 전체 단어 수준에서 활성화된 음절 이웃 단어의 방해로 알려져 있으며 이 방해의 크기는 표적 단어가 얼마나 많은 음절 이웃 단어를 또는 얼마나 강력한 음절 이웃 단어를 가지고 있는지에 의해 결정된다. 그러나 음절 빈도의 정의가 음절 타입 빈도와 토큰 빈도로 구분됨에도 불구하고 이를 구분하지 않고 많은 연구들이 수행되어 왔다. 최근 Conrad, Carreiras, & Jacobs(2008)에 따르면 음절 토큰 빈도는 전체 단어 처리 수준을 반영하는 변인이며 음절 타입 빈도는 하위 어휘 처리 수준의 음절 처리 수준을 반영하는 변인일 수 있다고 주장하였다. 이에 본 연구는 이들의 주장이 맞다면 음절 타입 빈도는 단어 명명 속도를 촉진 시킬 것이며 반대로 음절 토큰 빈도는 명명 시간과 관련 없을 것이라고 예측하였다. 왜냐하면 표기 심도가 얇고 음절의 경계가 명확한 언어에서 명명 과제는 전체 단어 수준을 덜 참고하기 때문이었다. 실험 1결과에서 음절 토큰 빈도를 통제한 상태에서 고빈도 타입 음절의 단어 명명 시간은 유의미하게 짧았다. 실험 2에서 음절 타입 빈도를 통제한 상태에서 음절 토큰 빈도의 증가는 명명 시간을 역시 단축시켰다. 이에 본 연구는 음절 토큰 빈도가 하위 어휘 처리와 무관하다는 Conrad, Carreiras, & Jacobs(2008)의 주장을 반박하였다.

주제어 : 음절 빈도, 음절 토큰 빈도, 음절 타입 빈도, 명명 과제

* 본 연구는 건국대학교 KU연구전임 프로그램에 의해서 수행된 과제임.

†교신저자: 권유안, 건국대학교 다언어 다문화 연구소, 서울특별시 광진구 능동로 120

연구 분야: 심리학, 언어심리학, Tel: 02-450-3944, E-mail: yakwon75@gmail.com

서 론

과거 시각 단어 재인 연구에서 단어 안에서 두 낱자의 출현 빈도(bigram frequency)가 상대적으로 낮은 지점이 음절의 경계라는 두 철자 골(bigram trough) 가설이 영어를 중심으로 지지되었었다[1]. 그러나 [2]은 스페인어를 이용하여 음절 경계의 두 낱자 빈도를 통제했음에도 불구하고 음절 빈도 효과가 나타남을 보고하였다. 음절 빈도 효과(syllable frequency effect)란 자주 사용하는 음절(이하, 고빈도 음절)로 시작되는 단어가 자주 사용하지 않는 음절(이하, 저빈도 음절)로 시작되는 단어에 비해 어휘 판단 시간이 느려지고 어휘 판단 정확도가 떨어지는 것을 의미한다. 이 연구 이후 음절 빈도 효과는 한국어[3]를 비롯하여 독일어[4], 프랑스어[5]에서도 동일한 효과가 일관되게 관찰되었고, 최근 반응 시간 결과뿐만 아니라 사건 관련 전위 연구에서도 음절 빈도 효과가 한국어[6], 독일어[7], 스페인어[8] 모두에서 동일한 유형으로 관찰되었다. 이에 연구자들은 과거 “표기 처리-전체 단어 처리 수준”만 가정하던 상호 활성화 모형(interactive activation model)에 음절 처리 단위를 함께 가정해야한다는 주장을 하였고, 이 주장은 음절의 경계가 영어에 비해 상대적으로 명확한 언어인 한국어, 독일어, 스페인어, 그리고 프랑스어에서 받아들여지고 있다[3, 5, 9, 10].

이론적으로 음절 처리 단위를 지지하는 연구자들은 상호 활성화 이론 중 하나인 Grainger와 Jacobs[11]의 MROM(multiple read-out model)을 음절 단위로 확장시켜 음절 빈도 효과를 설명하고 있다. MROM의 기본적 구조는 “낱자 처리 수준-자소 대 음소 대응 처리 수준-전체 단어 처리 수준”이며, 각 단계들은 축진적 연결이 되어 있지만 각 수준 안 표상들 간에는 억제적 연결이 되어 있다고 가정한다. 그리고 각 표상들의 활성화 역치는 자주 사용하는 표상인 경우 낮은 활성화 역치값을 갖고 있고, 반면 자주 사용하지 않는 표상인 경우 높은 역치값을 갖고 있다고 가정한다. 그러나 음절 빈도 효과를 지지하는 연구들은 두 철자 빈도가 음절 처리와 무관하다는 증거[2]를 바탕으로 자소 대 음소 대응 처리 수준 다음에 음절 처리 수준가 하나 더 있다고 주장한다[10]. 따라서 최근 모형은 MROM의 처리 가정을 유지한 채 “낱자 처리 수준-자소 대 음소 대응 처리 수준-음절 처리 수준-전체 단어 처리 수준”을 가정한다. 이 모형을 바탕으로 음절 빈도 효과

유발 과정을 다음과 같이 설명할 수 있다. 시각적으로 제시된 2음절 표적 단어의 낱자들은 자소 대 음소 처리를 통해 표적 단어와 음소를 공유하는 음절 처리 수준의 음절 표상을 활성화 시키게 된다. 이때 표적 단어의 첫 음절은 즉각적으로 처리되어 전체 단어 수준에 표상된 동일 음절로 시작되는 단어(이하 음절 이웃 단어)들을 활성화시킨다고 가정한다. 다음으로 전체 단어 수준에서 활성화된 음절 이웃 단어들은 서로 억제적 연결이 되어 있어 서로 억제적 활성화를 보내고 이 경쟁이 음절 빈도 효과의 원인이라고 설명한다[10]. 일반적으로 고빈도 음절로 시작되는 단어가 저빈도 음절로 시작되는 단어에 비해 음절 이웃 단어를 더 많은 갖고 있어 저빈도 음절의 단어에 비해 더 많은 경쟁을 유발하고 따라서 반응의 역제가 더 크게 발생한다고 보고 있다[2, 3].

그러나 흥미로운 점은 음절 처리 수준을 주장하는 연구들 안에서도 음절 빈도에 대한 정의를 다르게 했음에도 불구하고 수행된 연구들 간에 동일한 음절 빈도 효과가 관찰된다는 점이다. 대표적으로 [2]은 음절 빈도를 표적 단어와 같은 첫 음절을 공유하는 음절 이웃 단어의 개수로 정의하였고, [12]는 음절 빈도를 표적 단어와 첫 음절을 공유하는 음절 이웃 단어들의 출현 빈도의 합으로 정의하였다. 전자를 음절 타입 빈도라 하며 후자를 음절 토큰 빈도라 명명하는데, 두 연구에서 모두 그 빈도가 높아질수록 어휘 판단 반응이 억제되는 결과가 나타났다. 이러한 이유로 [13]은 서로 다른 음절 빈도의 정의 중 어떤 정의가 음절 빈도 효과를 반영하는지에 대한 질문을 제기 하였다.

[13]은 이 물음을 풀기 위해 음절 타입 빈도를 고정된 채 음절 토큰 빈도를 고저로 조작하여 어휘 판단 과제를 실시하였고, 추가적인 어휘 판단 과제에서는 반대로 음절 토큰 빈도를 고정된 채로 음절 타입 빈도 고저 간 차이가 어휘 판단 시간에 어떤 영향을 끼치는 지를 실험하였다. 결과에서 음절 토큰 빈도의 증가는 어휘 판단 시간의 증가와 관련 있었지만, 음절 타입 빈도의 증가는 반대로 반응 시간의 감소와 관련 있었다. 이 결과를 바탕으로 [13]은 논의에서 음절 타입 빈도와 토큰 빈도가 음절을 포함한 상호 활성화 모형에서 다른 처리 수준을 반영하기 때문에 서로 다른 효과가 나타날 수 있다고 주장하였다. 구체적으로 [13]은 MROM의 처리 과정과 같이 음절 처리 수준 안에는 음절 표상이 있고 이 음절 표상의 활성화 역치는 고빈도 음절인 경우 낮게 설정되어 있고, 저빈도 음절인 경우 높게

설정되어 있다고 주장하였다[12, 14]. 따라서 만약 어떤 음절이 그 언어에서 자주 사용되는 음절이라면 음절 처리 수준에 표상된 그 고빈도 음절은 자주 사용하지 않는 저빈도 음절에 비해 더 빨리 활성화되어 다음 처리의 단계로 처리 결과를 전달할 수 있다. 이와 같은 음절 처리 단계에서 고빈도 음절은 그 음절을 포함한 단어 자체의 빈도가 아니라 그 음절을 포함한 단어가 그 언어에서 얼마나 많이 있는가, 즉 음절 타입 빈도를 의미한다. [13]은 이와 같은 의미에서 고빈도 타입 음절이 저빈도 타입 음절보다 더 전형적 또는 친숙한 음절이라고 표현하였다. 다음 단계인 전체 단어 수준에는 단어 표상이 있고 이 단어 표상의 활성화 역치는 음절 표상과 마찬가지로 그 사용 빈도와 역함수관계에 있다고 가정한다. 따라서 고빈도 단어가 저빈도 단어에 비해 활성화 역치 수준이 낮아 더 빨리 활성화된다[5, 12]. 이 단계에서 단어의 활성화는 음절 타입 빈도가 아닌 단어 자체의 출현 빈도에 의해 결정되고 이 단어 빈도와 직접적 관련이 있는 음절 빈도가 바로 음절 토큰 빈도라고 [13]은 주장한다. 왜냐하면 음절 토큰 빈도의 증가는 음절 이웃 개수보다는 그 음절 이웃 단어들 중 단어 빈도가 높은 이웃 단어에 의해 더 큰 영향을 받기 때문이다. 예를 들어 한국어 단어에서 “억측”의 “억-”과 “낭송”의 “낭-”의 타입 빈도는 약 79로 동일하다. 그러나 “억-”의 토큰 빈도는 약 5512인 반면 “낭-”의 토큰 빈도는 약 1856이다. “억-”의 토큰 빈도가 높은 이유는 “억울: 단어 빈도 660”, “억압: 단어 빈도 605”와 같이 고빈도의 음절 이웃 단어가 있기 때문이다. 위 음절 빈도 효과에 대한 이론적 설명에서 음절 빈도 효과는 전체 단어 수준에서 활성화된 단어 표상들 간의 경쟁으로 인해 발생한다고 하였다. 이 경쟁을 일으키는 경쟁자가 표적 단어 보다 더 자주 출현하는 고빈도 단어인 경우 전체 단어 수준에서 활성화 역치가 표적 단어에 비해 더 낮아서 더 빠르게 활성화된다. 그리고 이 고빈도 이웃 단어가 표적 단어의 강력한 경쟁자가 되어 표적 단어 재인을 억제하게 된다.

요약하면 음절 타입 빈도는 어휘 이하 수준의 음절 표상을 얼마나 자주 쓰는가와 관련 있기 때문에 어휘 이하 음절 처리를 반영하고, 음절 토큰 빈도는 음절 처리 수준이 아닌 전체 단어 처리 수준에 표상된 단어의 빈도와 관련 있는 것으로 써, 결국 음절 토큰 빈도는 전체 단어 수준에 민감하게 영향을 끼친다는 것이 [13]의 논리이다.

이에 본 연구는 [13]의 주장을 검증하기 위해 음절 타입 및 토큰 빈도를 조작한 단어를 대상으로 단어 명명 과제를 실시하였다. 어휘 판단 과제가 아닌 단어 명명 과제를 사용한 이유는 단어 명명 과제가 전체 단어 수준의 처리를 어휘 판단 과제에 비해 덜 요구하기 때문이다. 명명 과제는 단어 자극 제시 후 즉각적으로 단어를 따라 말하도록 지시한다. 그래서 제시된 단어 자극이 어떤 의미인지 그리고 그 어휘성이 있는지 없는지 판단시키는 어휘 판단 과제에 비해 단어 명명 과제는 전체 단어 수준의 참고를 덜 요구한다[15]. 특히 표기 심도가 얇은 언어의 경우 표기 심도가 깊은 언어에 비해 전체 단어 수준의 정보를 참고하지 않는 경향이 더 강하다는 증거가 있다[15, 16]. 따라서 표기 심도가 얇은 한국어 단어 명명 과제를 사용하면 음절 타입 빈도가 어휘 이하 수준의 음절 처리를 반영하고, 음절 토큰 빈도가 전체 단어 처리 수준을 각각 반영한다는 [13]의 주장을 검증할 수 있을 것이다. 구체적으로 만약 음절 타입 빈도가 어휘 이하 처리인 음절 처리에 민감한 변인이라면, 고빈도 음절 타입 빈도로 시작되는 단어가 저빈도 음절 타입 빈도로 시작되는 단어에 비해 명명 시간이 짧을 것이다. 두 번째로 만약 음절 토큰 빈도가 전체 단어 처리 수준에 민감한 변인이라면 음절 토큰 빈도의 차이는 단어 명명 시간에서 나타나지 않아야 한다. 본 연구는 위와 같은 가설을 검증하여 [13]의 주장이 맞는지 아닌지를 검토하였다.

실험 1

실험 1은 고빈도 음절 타입으로 시작되는 단어가 저빈도 음절 타입으로 시작되는 단어에 비해 단어 명명 시간이 더 짧은지 아닌지를 검증하기 위해 음절 토큰 빈도를 고정한 상태에서 고빈도 타입 음절의 단어와 저빈도 타입 음절의 단어 간의 단어 명명 시간을 비교하였다. 만약 음절 타입 빈도가 음절 처리 수준에 민감하다면 고빈도 타입 음절의 단어가 저빈도 타입 음절의 단어에 비해 명명 속도가 더 빠를 것으로 예측되었다.

참가자

건국대학교에 재학 중인 30명(남: 17, 여: 13)을 모집하였다. 모두 한국어 모국어 화자였으며 평균 연령은 22.6세였다. 실험 참여에 대한 보상으로 5,000원을 제공 받았다.

실험 재료

실험에 사용된 자극은 모두 2음절 명사 단어들로 고빈도 타입 음절 조건 및 저빈도 타입 음절조건에 각각 16개씩 할당하였다¹⁾. 자극 선별에서 자극 단어는 1500만 어절의 코퍼스에서 약 10에서 30회 출현하는 저빈도 단어들 중에서 추출되었다. 저빈도 단어만을 선별한 이유는 음절 빈도 효과는 대부분 저빈도 단어들에서 나타나기 때문이었다[2-5]. 실험 자극의 대표적 예로 “탐지(출현 빈도: 28)”은 “탐-”으로 시작되는 단어가 약 131개로 고빈도 타입 음절 목록에 속하며, “묵상(출현 빈도: 25)”은 “묵-”으로 시작되는 단어가 약 86개로 저빈도 타입 음절 목록으로 할당되었다. 그리고 음절 타입 빈도 외에 단어 출현 빈도, 철자 이웃 크기(한 철자만 다르고 나머지 철자는 같은 단어 개수), 고빈도 철자 이웃 크기(철자 이웃 단어 중 표적 단어 보다 단어 빈도가 고빈도인 철자 이웃의 개수), 세 철자 빈도(세 철자가 연속적으로 나타나는 회수), 두 번째 음절 타입 및 토큰 빈도 그리고 주요 통제 요인인 첫 음절 토큰 빈도를 통제하였다(표 1 참고). 자극 속성들에 대한 통계 분석에서 음절 타입 빈도 간의 차이는 유의미하였지만($R(1, 30) = 10.11, p < .01$), 음절 토큰 빈도는 조건 간에 차이는 유의미하지 않았다($R(1, 30) = 2.46, p = .12$). 또한 다른 속성들도 조건 간 차이가 유의미하지 않았다($ps > .1$). 그리고 실험 자극에 대한 피험자들의 책략을 막기 위해 단어 빈도 20에서 50사이의 2음절 명사 64개를 무작위로 선택하여 명명 목록에 추가하였다. 사용된 단어 자극은 모두 21세기 세종계획에서 배포한 1500만 어절 코퍼스에서 선별되었다[17].

1) 실험 자극의 개수가 다른 연구에 비해 적은 이유는 단어 출현 빈도, 철자 이웃 크기, 고빈도 철자 이웃 크기, 세 철자 빈도, 두 번째 음절 타입 및 토큰 빈도 등을 모두 통제했기 때문이다.

표 1. 실험 1 자극의 속성, 평균값(표준편차)

자극 속성	고빈도 음절 타입 자극	저빈도 음절 타입 자극
첫 음절 타입 빈도	5.0 (0.3)	4.4 (0.1)
첫 음절 토큰 빈도	8.1 (0.4)	7.9 (0.4)
단어 출현 빈도	2.9 (0.7)	2.9 (0.6)
고빈도 철자 이웃 개수	0.4 (0.3)	0.2 (0.3)
철자 이웃 개수	1.3 (0.7)	1.0 (0.7)
세 철자 빈도	8.4 (0.7)	8.7 (0.7)
두 번째 음절 타입 빈도	5.9 (0.6)	5.8 (1.0)
두 번째 음절 토큰 빈도	10.3 (0.8)	10.2 (1.5)

참고) 빈도값은 모두 원자료 값을 log10으로 치환한 값이고 개수 값은 원자료 값이다.

절차

실험 참가자가 자극이 제시되는 모니터 앞 약 70-90cm 앞에 앉으면, 실험자는 준비된 실험 지시 사항을 읽어 주었다. 실험 지시문은 화면 중앙에 제시되는 단어 자극을 가급적이면 빠르고 정확하게 읽으라는 내용과 마이크에 읽기 외의 소리를 내지 말 것을 주의 주는 내용이었다. 단어 명명 과제의 절차는 자극 응시점(+)이 약 1초간 화면 중앙에 제시되고 이것이 사라지자마자 자극 단어가 나타나며 이 자극은 명명을 하지 않을 경우 약 2초 동안 자극이 유지되고 사라진 후 약 1초의 수행 간 공백이 제시 되었다. 하지만 참가자가 명명을 하면 자극 단어는 즉각 사라지고 수행 간 공백이 제시되었다. 자극은 검은 바탕에 흰 글씨로 제시되었으며 글자 크기는 26폰트였다. 실험 참가자는 15회의 연습 시행을 한 후 본 시행을 수행하였다. 만약 실험 참가자가 과제를 이해하지 못한 경우 연습 시행을 다시 반복시켰다. 모든 수행은 디지털 녹음기로 기록되었고 이 기록을 바탕으로 정오 판정을 수행하였다. 정오 판정의 기준은 틀린 발음, 명명 이외의 소리(예, ‘쩍쩍’ 소리, 기침 소리, ‘음’하는 소리 등) 등은 오답으로 기록하였다.

실험 도구

자극들은 모두 17" CRT 모니터 중앙에 제시되었으며 명명 반응은 Psychology software tools 회사의 serial response box에 연결된 목소리 반응 마이크로 받았다. 자극의 제시 및 기록은 같은 회사의 e-prime 2 version을 사용하였다.

실험 1 결과

오류를 많이 범한 실험 참가자 4명(>15%)을 분석에서 제외하였다. 단어 명명 시간에 분석은 자극 무선 요인과 피험자 무선 요인을 동시에 고려하는 선형 혼합 효과 모형(linear-mixed effects) 통계 분석을 실시하였다[18]. 오류율 분석은 전통적인 $F_1 \times F_2$ 분석을 사용하였다²⁾. 표 2에 26명의 평균 명명 시간과 오류율을 제시하였다.

표 2. 실험 1 평균 명명 시간과 오류율, 괄호 안은 표준 편차

	고빈도 타입 음절	저빈도 타입 음절
평균 명명 시간(ms)	424 (43)	436 (50)
평균 오류율(%)	5.1 (1.3)	4.6 (1.2)

분석 결과 고빈도 타입 음절로 시작되는 단어의 명명 시간이 저빈도 타입 음절로 시작되는 단어에 비해 더 빨랐다. 그리고 이 차이는 통계적으로 유의미하였다 ($t(790.90) = 2.42, SE = 5.15, p < .05$)[20]. 그러나 오류율의 차이는 없었다($F(1, 25) = 2.39, p = .13; F(1, 30) = .19, p = .66$). 실험 결과 [13]의 주장대로 고빈도 음절 타입 빈도로 시작되는 단어가 저빈도 타입 빈도로 시작되는 단어에 비해 더 빠르게 나타났다.

2) 오류율 분석에 선형 혼합 모형 분석을 사용하지 않은 이유는 원자료가 1또는 0의 범주적 자료였고, 본 연구에서 평균 오류율이 주요 종속변인이 아니었기 때문에 복잡한 분석을 실시하지 않았다.

실험 2

실험 2는 실험 1의 조건과 반대로 음절 타입 빈도를 고정하고 음절 토큰 빈도를 고빈도와 저빈도로 나누어 명명 과제를 실시하였다. 만약 [13]의 주장처럼 음절 토큰 빈도가 전체 단어 처리 수준에 민감한 변인이라면 전체 단어 수준의 정보를 상대적으로 덜 참고하는 명명 과제에서 단어 명명 속도는 음절 토큰 빈도의 변동에 영향을 거의 받지 않을 것으로 예측되었다.

참가자

건국대학교에 재학 중인 26명(남: 12, 여: 14)을 모집하였다. 모두 한국어 모국어자였으며 평균 연령은 21.6세였다. 실험 참가비로 5,000원이 참가자에게 지급되었다.

실험 재료

실험에 사용된 단어 자극은 고빈도 토큰 음절 단어 18개 저빈도 토큰 음절 단어 18개를 실험 1과 마찬가지로 21세기 세종계획 코퍼스에서 선택하였다. 자극 단어 예를 들면, “담소(출현 빈도: 27)”의 “담-”으로 시작되는 단어 약 196개의 출현 빈도의 총합은 11997이었고 “탐지(출현 빈도: 28)”의 “탐-”으로 시작되는 단어 131개의 출현 빈도의 총합은 2642였다. 실험 1과 같이 단어 출현 빈도, 고빈도 철자 이웃 크기, 철자 이웃 크기, 세 철자 빈도, 두 번째 음절 타입 및 토큰 빈도 그리고 주요 통제 요인인 첫 음절 타입 빈도를 통제하였다(표 3 참고). 통계적 분석에서 토큰 음절 빈도 간의 차이는 유의미 하였으나($F(1, 35) = 13.06$, $p < .001$), 음절 타입 빈도는 고빈도 음절 토큰과 저빈도 음절 토큰 간에 유의미한 차이를 보이지 않았다($F(1, 35) = .61$, $p = .44$). 마찬가지로 다른 통제 변인들도 유의미한 차이를 보이지 않았다($ps > .1$). 채움 단어 목록은 실험 1과 동일하였다.

표 3. 실험 2 자극의 속성, 평균값(표준편차)

자극 속성	고빈도 음절 토큰 자극	저빈도 음절 토큰 자극
첫 음절 토큰 빈도	8.7 (0.6)	7.6 (0.1)
첫 음절 타입 빈도	4.8 (0.2)	4.7 (0.2)
출현 빈도	3.1 (0.6)	2.8 (0.7)
고빈도 철자 이웃 개수	0.4 (0.3)	0.3 (0.3)
철자 이웃 개수	1.3 (0.7)	1.1 (0.7)
세 철자 빈도	8.9 (0.9)	8.4 (0.8)
두 번째 음절 타입 빈도	4.8 (0.2)	4.7 (0.2)
두 번째 음절 토큰 빈도	8.7 (0.6)	7.6 (0.1)

참고) 빈도 값은 모두 원자료 값을 log10으로 치환한 값이고 개수 값은 원자료 값이다.

절차 및 실험 도구

실험 2의 절차 및 실험 도구는 모두 실험 1과 동일하였다.

실험 2 결과

실험 참가자의 손실은 없었다. 실험 1과 동일한 분석 방법을 사용하였다. 분석 결과 고빈도 음절 토큰이 저빈도 음절 토큰에 비해 명명 시간이 더 빠르게 나타났다($t(892.80) = 4.45, SE = 4.77, p < .001$). 오류율 분석에서 두 조건 간의 차이는 없었다($F(1, 25) = 2.39, p = .13; F(1, 34) = .29, p = .59$). 표 4에 평균 명명 시간 및 오류율을 제시하였다.

표 4. 실험 2 평균 명명 시간과 오류율, 괄호 안은 표준 편차

	고빈도 토큰 음절	저빈도 토큰 음절
평균 명명 시간(ms)	416 (43)	434 (49)
평균 오류율(%)	5.3 (1.4)	4.9 (1.2)

분석 결과 [13]의 주장과는 다르게 고빈도 토큰 음절로 시작되는 단어가 저빈도 토큰 음절로 시작되는 단어에 비해 명명 시간이 더 빨랐다. 이는 음절 토큰 빈도 역시 어휘 이하 음절 처리에 관련 있음을 시사한다.

종합 논의

본 연구는 음절 빈도가 음절 토큰 빈도와 음절 타입 빈도로 나뉘며 음절을 포함한 상호 활성화 모형에서 음절 타입 빈도는 어휘 이하 단계인 음절 처리 수준을 반영하고, 음절 토큰 빈도는 전체 단어 처리 수준을 반영한다는 [13]의 주장을 단어 명명 과제를 통해 검증하였다. 실험 1 결과, 단어 명명 시간은 고빈도 타입 음절로 시작하는 단어가 저빈도 타입 음절로 시작하는 단어에 비해 더 빨랐다. 실험 2 결과에서도 고빈도 토큰 음절의 단어가 저빈도 토큰 음절의 단어에 비해 단어 명명 시간이 더 빨랐다. 실험 1의 결과는 음절 타입 빈도에 대한 [13]의 주장과 일치하지만 음절 토큰 빈도에 대한 [13]의 주장은 기각되었다. 이러한 이유로 음절 토큰 빈도에 대한 [13]이 주장을 보다 자세히 논의할 것이다.

본 연구에서 관찰된 음절 토큰 빈도 증가에 따른 명명 시간의 축진은 비단 한국어에서만 나타난 결과가 아니다. 스페인어와 네덜란드어에서도 동일한 효과가 보고되었다[15, 21, 22]. [15]의 스페인어 연구는 고빈도 토큰 음절로 시작되는 단어가 저빈도 토큰 음절로 시작되는 단어에 비해 명명 시간이 더 빠른 것을 관찰하였다. 스페인어 결과에서 보다 중요한 결과는 사후 분석에서 음절 토큰 빈도 증가에 따른 명명 시간 축진 효과가 고빈도 음절 이웃 개수와 음절 타입 빈도와 무관했다는 점이다. 이 의미는 전체 단어 처리 수준의 억제적 활성화 정도를 반영하는 고빈도 음절 이웃의 개수와 명명 시간이 관련 없다는 것을 의미한다. 이러한 결과는 네덜란드어 연구인 [21, 22]에서도 관찰되었다. 네덜란드어 연구 결과도 음절 토큰 빈도가 증가할수록 명명 시간이 빨라지는 것을 관찰하였다. 이들의 설명에 따르면 음절은 음절 체계(mental syllabary)로 저장되어있어 자주 사용하는 음절일수록 음절 체계(mental syllabary)에 빠르게 접속되고 처리되어 운동 출력(motor output)에 더 빨리 전달되기 때문에 명명 시간이 빨라진다고 설명하였다. 이 결과들을 종합하면,

명명 과제에서 고빈도 토큰 음절의 촉진적 효과가 전체 단어 수준이 아니라 하위 어휘 단계의 음절 처리 수준과 더 밀접한 관련이 있어 보인다.

부가적으로 음절 타입 빈도 증가로 인한 명명 시간의 촉진도 여러 연구들의 결과와 일치한다. 특히 프랑스어 연구인 [23]에서 고빈도 타입 음절인 경우가 저빈도인 경우에 비해 명명 시간이 더 빨랐다. 이에 [23]은 음운 어휘 이하 입력 수준(input sublexical phonology)과 음운 출력 어휘 이하 수준(output sublexical phonology) 사이에 음절 단위가 표상되어 있고 이 음절 단위 표상은 고빈도 음절인 경우 음운 어휘 이하 수준 출력(output sublexical phonology)을 더 빨리 활성화시켜 명명 시간이 촉진되었다고 설명하였다. 그리고 이 설명은 영어 연구에서도 적용되었다[24]. 따라서 음절 타입 빈도가 어휘 이하 음절 표상의 처리를 반영한다는 [13]의 주장은 무리가 없어 보인다.

만약 음절 토큰 빈도가 [13]의 주장처럼 전체 단어 처리 수준의 역제적 활성화와 관련있는 것이라면, 본 연구를 비롯한 [15, 21-24]의 결과와 다르게 음절 토큰 빈도의 증가가 명명 시간의 촉진을 유발해서는 안된다. 하지만 음절 토큰 빈도의 증가가 단어 명명 시간을 일관되게 촉진시켰다는 것은 음절 토큰 빈도와 음절 타입 빈도가 [13]의 주장과 달리 독립적이지 않다는 것을 의미한다. 실제로 [13]도 인정하였듯이 전체 코퍼스에서 음절 타입 빈도와 토큰 빈도의 상관은 매우 높은 편이다. 본 연구가 사용한 전체 코퍼스($n=192,300$)에서도 음절 타입 빈도와 음절 토큰 빈도는 높은 정적 상관($r=.66, p < .001$)을 보였다. 이 의미는 실제 언어활동에서 두 빈도가 서로 독립적이 않을 수 있다는 것을 의미한다.

그리고 [13]은 사건 관련 뇌파 연구에서 음절 토큰 빈도의 증가에 따라 N400의 평균 진폭이 증가하는 점을 근거로 음절 토큰 빈도가 전체 단어 처리 수준을 반영한다고 주장하였다. [13]의 사건 관련 뇌파 연구인 [7]은 음절 토큰 빈도가 증가하는 경우 단어의 의미 처리를 반영하는 N400 평균 진폭의 크기가 커지는 것을 관찰하였다. 그리고 설명에서 [25]의 연구를 바탕으로 고빈도 음절 이웃 단어가 활성화시킨 의미와 표적 단어가 활성화시킨 의미 간의 경쟁이 발생하고, 이 경쟁은 처리의 부담을 일으켜 N400 평균 진폭의 크기가 다르게 나타났다고 주장하였다. 그러나 [13]이 인용한 [25]의 연구는 표기 이웃 빈도 효과(orthographic neighborhood frequency effect)가 아닌 표기 이웃 크기 효과(orthographic neighborhood density effect),

즉 타입 빈도의 효과를 검증한 것이었다. 또한 [25]는 표기 이웃이 많은 표적 단어가 표기 이웃이 적은 표적 단어에 비해 더 많은 단어 표상을 활성화시키고, 이 단어들과 연합된 의미들이 즉각적으로 더 많이 활성화되어 N400 평균 진폭이 커진다고 설명하였다. 이를 음절 단위로 적용하면 고빈도 음절 이웃 단어와 연합된 의미가 아닌 같은 음절을 공유하는 음절 이웃 단어들과 연합된 의미가 많을수록 N400 평균 진폭이 커진다는 것이다. 이를 [13]은 확대 해석한 것으로 보인다. 무엇보다도 보다 강력한 반대 증거는 한국어 결과에서 제시되었다. [14]는 의미 범주 판단 과제에서 고빈도 표기 이웃 개수를 통제했을 때 N400 평균 진폭의 크기가 음절 타입 빈도의 증가와 밀접히 관련 있다는 것을 보여주었다. 따라서 음절 타입 빈도와 음절 토큰 빈도가 수치적으로 구분되고 개념적으로 구분된다 할지라도 실제로 각 빈도가 처리의 수준을 독립적으로 반영한다는 경험적 증거는 아직 부족하다고 할 수 있다.

마지막으로 음절을 포함한 상호 활성화 모형을 본 연구 결과에 적용하면 다음과 같다. [13]은 어휘 이하 수준의 음절 표상들의 활성화 역치값이 음절 타입 빈도에 의해 결정된다고 보았지만, 본 연구 결과와 이를 지지하는 여러 연구들을 종합하면 음절 표상의 역치값은 음절의 타입과 토큰 빈도와 모두 관련 있는 것으로 보인다. 음절 토큰 빈도가 높다는 것은 분명 그 언어에서 자주 사용하는 단어가 그 음절을 포함하고 있다는 것을 의미한다. 따라서 고빈도 토큰 음절 표상은 저빈도 토큰 음절 표상보다 더 자주 사용되었을 것이며, 결국 고빈도 토큰 음절 표상의 활성화 역치값은 저빈도의 그것에 비해 저 낮게 설정되었을 수 있다. 이 같은 맥락에서 전체 단어 처리 수준을 주로 참고하지 않는 명명 과제에서 고빈도 토큰 음절의 단어가 저빈도에 비해 더 빠른 명명 시간이 나타난 것을 설명할 수 있다.

결론적으로 [13]에서 어휘 판단 과제가 명명 과제에 비해 전체 단어 수준에 더 민감해서 음절 토큰 빈도의 영향이 나타난 것이지 음절 토큰 빈도가 하위 어휘 수준의 음절 표상에 영향을 주지 않았기 때문은 아닌 것으로 보인다.

참고문헌

- [1] Seidenberg, M., & McClelland, J. L. (1989). A distributed, developmental model of visual word recognition and naming. *Psychological Review*, 96, 4, 523-568.
- [2] Carreiras, M., Alvarez, J. C., & De Vega, M. (1993). Syllable frequency and visual word recognition in Spanish. *Journal of Memory and Language*, 32, 766-780.
- [3] Kwon, Y., Lee, C., 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.
- [4] 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.
- [5] Mathey, S., Zagar, D., Doignon, N., & Seigneuric, A. (2006). The nature of the syllabic neighborhoods effect in French. *Acta Psychologica*, 123, 372-393.
- [6] 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.
- [7] 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.
- [8] Barber, H., Vergara, M., & Carreiras, M. (2004). Syllable-frequency effects in visual word recognition: Evidence from ERPs. *Neuroreport*, 15(3), 545-548.
- [9] Conrad, M., Grainger, J., & Jacobs, A. M. (2007). Phonology as the source of syllable frequency effects in visual word recognition: Evidence from French. *Memory and Cognition*, 35(5), 974-983.
- [10] Perea, M., & Carreiras, M. (1998). Effects of syllable frequency and syllable neighborhood frequency in visual word recognition. *Journal of Experimental Psychology: Human Perception & Performance*, 24, 134-144.
- [11] Grainger, J., & Jacobs, A. M. (1996). Orthographic processing in visual word

- recognition: A multiple read-out model. *Psychological Review*, 103, 518-565.
- [12] 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.
- [13] 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.
- [14] Kwon, Y., Nam, K., & Lee, Y. (2012). ERP index of the morphological family size effect during word recognition. *Neuropsychologia*, 50, 3385-3391.
- [15] Carreiras, M., & Perea, M. (2004). Naming pseudowords in Spanish: Effects of syllable frequency. *Brain & Language*, 90, 393-400.
- [16] Conrad, M., Stenneken, P., & Jacobs, A. M. (2006). Associated or dissociated effects of syllable frequency in lexical decision and naming. *Psychonomic Bulletin & Review*, 13(2), 339-345.
- [17] 김흥규, 강범모. (2008). **한국어 단어와 형태소의 사용 빈도 1500만 어절의 세종형태 의미 분석 말뭉치 기반**. 서울: 고려대학교 민족문화연구원.
- [18] Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59, 390-412.
- [20] Winter, B. (2013). Linear models and linear mixed effects models in R with linguistic applications. arXiv:1308.5499. [http://arxiv.org/pdf/1308.5499.pdf]
- [21] Cholin, J., Schiller, N. O., & Levelt, W. J. M. (2004). The preparation of syllables in speech production. *Journal of Memory and Language*, 50, 47-61.
- [22] Cholin, J., Levelt, W. J. M., & Schiller, N. O. (2006). Effects of syllable frequency in speech production. *Cognition*, 99, 205-235.
- [23] Ferrand, L., Segui, J., & Grainger, J. (1996). Masked priming of word and picture naming: the role of syllabic units. *Journal of Memory and Language*, 35, 708-723.
- [24] Cholin, J., Levelt, W. J. M., & Schiller, N. O. (2006). Effects of syllable frequency in

speech production. *Cognition*, 99, 205-235.

- [25] Holcomb, P. J., Grainger, J., & O'Rourke, T. (2002). An electrophysiological study of the effects of orthographic neighborhood size on printed word perception. *Journal of Cognitive Neuroscience*, 14(6), 938-950.

1 차원고접수 : 2014. 01. 30

2 차원고접수 : 2014. 03. 23

최종게재승인 : 2014. 04. 11

(*Abstract*)

The Syllable Type and Token Frequency Effect in Naming Task

Youan Kwon

Multilingualism & Multiculturalism Research Center, Konkuk University

The syllable frequency effect is defined as the inhibitory effect that words starting with high frequency syllable generate a longer lexical decision latency and a larger error rate than words starting with low frequency syllable do. Researchers agree that the reason of the inhibitory effect is the interference from syllable neighbors sharing a target's first syllable at the lexical level and the degree of the interference effect correlates with the number of syllable neighbors or stronger syllable neighbors which have a higher word frequency. However, although the syllable frequency can be classified as the syllable type and token frequency, previous studies in visual word recognition have used the syllable frequency without the classification. Recently Conrad, Carreiras, & Jacobs (2008) demonstrated that the syllable type frequency might reflect a sub-lexical processing level including matching from letters to syllables and the syllable token frequency might reflect competitions between a target and higher frequency words of syllable neighbors in the whole word lexical processing level. Therefore, the present study investigated their proposals using word naming tasks. Generally word naming tasks are more sensitive to sub-lexical processing. Thus, the present study expected a facilitative effect of high syllable type frequency and a null effect of high syllable token frequency. In Experiment 1, words starting with high syllable type frequency generated a faster naming latency than words starting with low syllable type frequency with holding syllable token frequency of them. In Experiment 2, high syllable token frequency also created a shorter naming time than low syllable token frequency with holding their syllable type frequency. For that reason, we rejected the propose of Conrad et al. and suggested that both type and token syllable frequency could relate to the sub-lexical processing.

Key words : syllable frequency effect, syllable type frequency, syllable token frequency, naming task