

한글 단어 인지과정에서
음운적 처리와 어휘근접

이 광오

영남대학교 문과대학 심리학과

A preliminary study on lexical access and
phonological processing in written word recognition

Kwangoh Yi

Department of psychology Yeungnam University

요 약

단어 인지과정은 언어 이해과정의 한 부분으로, 신속성과 정확성 그리고 심성어휘집을 그 특징으로 한다. 표기 단어의 인지과정에는 그 언어의 정서법 체계가 반영된다. 한국어의 단어 인지과정에 대한 모델 작성의 예비 연구로 음운적 처리와 어휘 근접에서의 정서법적 정보의 역할에 대해 검토하였다. 어휘 근접의 단위에 대한 논의에서는 음절, 자질, 단어등의 형식적 언어학적 단위외에 심리적 단위가 고려되어야 함이 지적되었으며, 그 심리적 단위들과 정서법의 관계에 대해 논의하였다. 마지막으로 한글 단어 인지과정에 관한 한 모델로서 상호작용 활성화 모델의 가능성에 주목하였다.

0.0 단어인지과정의 일반적특징

단어인지과정의 첫번째 특징은 신속성이다. 단어를 읽거나 듣고 의미를 이해하기까지 걸리는 시간은 문자 그대로 순식간이다. RSVP(Rapid Serial Visual Presentation) 하에서 1 초 동안에 20 개까지의 단어인지가 가능하며(Potter, 1984), 따라말하기(shadowing)과제에서는 적어도 150ms 이내에 음성단어의 인지가 가능하다(Marslen-Wilson, 1978). 그럼에도 불구하고 단어인지는 정확하다. 정상적인 상황에서 적당한 주의가 있으면 잘못 읽는 일은 없다. 또 한가지 특징은 단어인지과정이 방대한 데이터베이스를 취급한다는 것이다. 인간이 기억하고 있는 단어의 수는 적게잡아도 1-2 만 개이다(이것을, 인간의 마음속에 있는 사전이라하여, 심성어휘집 mental lexicon 이라 한다). 이 방대한 어휘집의 검색이 어떻게 신속하고 정확하게 이루어지는가를 설명하는 것이 단어인지과정 연구의 하나의 목적이다.

1.0 단어인지과정의 연구문제

단어인지과정이 신속하고 정확한 것은 심성어휘집이 단순한 저장고가 아님을 의미한다. 심성어휘집은 고도로 체계화되어 있는 데, 그 구조를 밝히는 것이 단어인지과정 연구의 한 과제이다. 또 한 가지 연구과제는 심성어휘집을 검색하여 심성어휘집에 있는 정보를 인출하는 과정에 대한 것이다. 제시된 단어를 인지하기 위해서는 심성어휘집의 음립말인 어휘항목을 검색하는 과정이 필요한 데, 그 과정을 어휘근접 (lexical access) 이라 한다. 그리고 최종적으로 해당 어휘항목이 가진 정보를 인출하게 되는 데, 여기까지를 단어인지과정이라 한다. 단어인지와 어휘접근이 동의어로 사용되는 일이 있는 것은, 어휘근접이 단어인지과정의 중심적인 과정이기 때문일 것이다. 어휘근접의 근접부호와 근접단위는 정서법 (orthography) 과 불가분의 관계에 있으므로, 언어마다 차이가 있을 것이다. 다음에서는 한글

단어 인지과정의 근접단위에 대해서 고찰한다.

1.1 어휘근접의 단위

어휘근접과정에서 단위가 될 만한 후보에는 단어, 형태소, 음절, 자모, 자질(feature)들이 있는 데, 이들 사이에는 다음과 같은 관계가 있다.

- 단어 = {형태소}
- 단어 = {음절}
- 단어 = {문자}
- 형태소 = {음절, 자모}
- 음절 = {자모}
- 음절 = {onset, rime}
- 자모 = {자질}

이러한 형식적 언어학적 단위들이 모두 심리과정인 어휘근접의 단위가 되느냐에 대해서는 논란이 많다. 특히 영어 단어인지과정의 연구에서는 처리단위로서의 음절의 존재에 대한 대립되는 견해들이 발표되어 있다. 이것은 아마 영어 단어안에서의 음절이 선명치 못하기 때문일 것이다. 이와는 달리 음절을 선명하게 표시하는 한글단어에서는 음절은 어휘근접의 단위로서 매우 가능성이 높은 존재이다. 언어학적 단위가 아닌 순 심리적 단위의 존재가능성도 또한 크다. 예를 들어 심리적으로는 단어의 선두요소들이 말미요소들보다 중요한 데 이것은 단어의 구성요소들이 모두 동일한 주의를 받는 것이 아님을 의미한다. 영어의 경우, 어휘근접의 단위로서 단어선두의 자음 + 모음 + 자음에 해당하는 BOSS(Basic Orthographic Syllabic Structure) 라는 단위가 제안되어 있다(Taft, 1979). 이 제안에 따르면, 제시된 단어에서 먼저 BOSS 를 발견하고, 이어서 BOSS 를 근접부호로 하여 어휘집을 검색한다. 한글단어에 대해서 BOSS 단위를 제안하는 것은 몇 가지 문제가 있는 데, 그것은 한글단어의 음절이 지나치게 선명하기 때문이다. 그러나, 선두의 음절이 말미의 음절과 어휘근접 과정에서의 역할이 동등하지 않다는 의미에서의 BOSS 가설은 한글 단어에 대해서도 오히려 불필요가 있다.

똑같은 지적을 음절의 구성요소들에 대해서도 할 수 있다. 즉 음절을 구성하는 자모의 역할이 모두 동일하지 않을 가능성에 대해서이다. 예를 들어, 초성자모와 종성자모의 역할은 여러 면에서 차이가 있다고 본다. 만약 음절내에서 시각집단화(perceptual grouping)가 이루어진다면 그것은 어떤 형태일까? 초성자모와 종성자모를 집단화하는 방법과 종성자모와 종성자모를 집단화하는 방법이 있겠다. 후자는 onset 과 rime 을 요소로 하여 음절을 분석하는 음성학적 분석이다. 전자에 대해서는 심리학적인 지지가 있다. Marslen-Wilson(1978)의 cohort 모델에 의하면, 음성언어의 단어인지과정에는 cohort 를 생성하는 단원적(modular)인 과정이 있는 데, 이 과정은 단어 선두의 150ms 정도의 음성정보를 이용하고 있다. 이 시간은 대략 자음 + 모음에 해당하는 시간인데, 이 시간에 해당하는 양의 음성정보가 어휘근접의 단위가 되는 것이다. 만약에 음성단어 인지와 표기단어의 인지가 대응한다면, 한글표기한 음절안에서 중요한 것은 초성자모와 종성자모이다. (여기서는, 이것을 음절핵이라고 부르기로 한다.) 한국어에는 크게 두종류의 음절이 있는 데 (CV 와 CVC), 위 고찰에 의하면, CV 형이 보다 더 기본적인 음절이라고 해야 할 것 같다.

cohort 모델을 한글 음절의 처리에 적용하면, 음절의 처리는 음절핵을 중심으로 이루어질 것이고, 음절핵의 처리는 그 음절핵을 공유하는 모든 음절을 활성화한다. 이어서 나머지 자모, 문맥 또는 자질등을 이용하여 선택하는 과정을 거쳐 음절이 인지되는 것으로 생각된다 (그림 1 참조). (참고로, 한국어의 음절핵은 모두 399 개이며, 종성자음은 27 개이다.)

그림 1.

1.1.1 자질

단어인지과정의 초기과정은 자질분석과정이다. 자질에는 크게 두 종류가 있다. 단어인지과정에서 대국적 (global) 자질은 단어 또는 음절의 전체적 형태와 관계가 있는 자질이다. 디국적 자질의 후보로는 명암의 본포, 운곽선의 모양등이 있다. 예를 들어, 운곽선의 돌출횟수와 돌출장소의 분석에 의해 얻어지는 정보를 사용하면, 음절 모양의 분류가 어느 정도 가능하다. 물론 대국적 자질의 분석만으로 단어나 음절의 인지가 성립하지는 않을 것이다. 오히려, 기하학적 선분분석을 주로하는 국소적 자질 분석과 대국적 자질 분석을 병행하는 것이 단어 인지의 시간을 단축하고 인지의 정확성을 높일 수 있다.

자질 분석 단계에서 짚고 넘어가야 할 것 중에 분석의 초점의 문제가 있다. 앞서서도 지적했듯이, 단어나 음절의 구성요소는 모두가 동일하게 취급되지는 않는 것 같으며, 좀 더 중요한 구성요소라고 부를 수 있는 것들이 있기 때문이다. 단어나 음절의 상하좌우 어느 곳에 있는 요소가 보다 더 중요한 역할을 하는가의 문제는 경험적 자료를 기다려야 하겠으나, 앞에서 언급했던 음절력의 개념으로 부터 연역하건대 위쪽과 오른쪽이 분석의 초점이 될 것으로 생각된다 (그림 2 참조).

1.1.3 음절

음절은 한글단어의 선명한 단위가기 때문에, 처리단위의 유력한 후보이다. 단독 제시된 음절을 읽는 방법은 사람마다 차이가 없으며, 그 음절이 단어안에서 어떻게 읽히는 가와도 원칙적으로 무관하다. 따라서 한글 음절을 읽는 과정에 대한 단편적인 연구가 가능하다. 단독제시된 음절의 읽기에는 두 가지 방법이 사용될 수 있다. 하나는 구성자모를 이용한 규칙에 의한 읽기이다. 이른바 영어의 GPC (Grapheme Phoneme Conversion) 규칙같은 것이 적용되는 읽기이다. 예를 들어, 단독제시된 음절 "강" "장" "값" "값" 등을 GPC 규칙을 적용하여 읽는 방법은, 자모 하나씩 소리로 바꾸어 그것을 연결하여 발음하는 것이다. 또 하나는 구성자모를 사용하지 않고, 음절의 전체적 (wholistic) 처리후에, 그 음절의 읽기정보를 인출하여 발음하는 경우이겠다. 이른바 어휘근접후 음운적정보의 인출과 닮은 과정이다. 전자는 한글

자모가 소리와 엄밀하게 대응하고 있는 사실을 잘 이용한 방법으로 상식과 일치한다. 반면 후자의 경우는 한글자모의 표음적 특징을 겁어든 채, 모양으로서만 이용하고 있어, 일견 비효율적일 것 같다.

그러나 한 가지 분명한 것은 단어속에 나타나는 음절에 대해서는 GPC 규칙에 의한 읽기가 적용되지 않는 예가 있다는 것이다. 예를 들어, "털보" 와 "들보" 의 제 2 음절은 표기적 동일성에도 불구하고 발음은 서로 다른 데 ("냉국" 과 "상국", "장독" 과 "낭독" 도 마찬가지다), 이것은 음절의 읽기가 자모수준에서만 결정되는 것이 아님을 시사한다. 이러한 예는 음운적 정보의 인출에 의한 읽기를 지지한다. 이 문제는 어휘근접이 음운을 매개로 하는가 아닌가 하는 문제와도 밀접한 관계가 있다.

1.1.4. 단어

단어 그 자체가 가진 속성으로 사용빈도가 있다. 사용빈도가 높은 단어가 사용빈도가 낮은 단어보다 빨리 인지된다는 사실은 영어를 비롯한 여러 언어에서 확인되어 있으며, 한국어에서도 또한 그렇다 (또는, 그럴 것이다). 단어의 길이, 단어의 형태소론적 (morphological) 인 구조등도 단어의 속성으로 단어인지과정에 중요한 요인들이다. 그러나 단어수준에서 무엇보다도 중요한 것은 단어와 단어사이의 관계일 것이다. 점화 (priming) 효과는 이 관계가 구체적으로 드러나는 경우이다. 단어사이의 관계에 대해서는 의미적 연상관계 (semantic association), 의미적 관련 (semantic relation) 등 다소 모호한 포괄적인 용어가 사용되고 있는데, 구체적인 하위 분석의 여지는 얼마든지 있다. 예를 들어, 연상관계로 지칭되는 단어간 관계에도, 전체-부분 관계, 대립관계, 유사관계, 전이관계 등등이 있으며, 모두 점화효과가 관찰되나 그 크기는 같지 않다. 어쨌든 단어들 사이에는 수많은 관계가 존재하며, 그것은 심성어휘집에 반영되어 있을 것이다. 점화효과와 같은 수준내 (intralevel) 문맥효과가 관찰되는 것은 바로 심성어휘집의 구조에 기인하는 것이다. 심성어휘집에 표현되어 있을 단어간 관계를 해명하는 작업이 앞으로 기대된다. 아울러, 단어의미구조의 해명, 단어의 문법속성의 기술, 그리고 단어사전으로서의

심성어휘집과 세상지 (world knowledge) 를 반영하는 개념사전과의 관계의 해명등도 연구되어야 할 과제들이다.

2.0 단어 인지 과정의 모델

지금까지 경험적 자료들이 충분히 축적되지 않은 상태에서, 한글 단어 인지과정의 연구에서 제기될 수 있는 몇 가지 문제들을 논의하였다. 아직 한글 단어인지 과정에 대한 모델을 제시할 시기는 아니므로, 모델 구성시에 고려해야 할 몇 가지 제한에 대해 살펴보겠다.

단어인지 과정은 언어이해 과정의 한 부분이다. 따라서, 단어인지 과정의 모델은 언어이해 과정의 모델과 밀접한 관계에 있어야 한다. 언어 이해과정의 한 부분으로서 단어인지 과정의 단원성 (modularity) 은 이미 여러 선행연구들이 지적하고 있으므로, 모델은 우선 이 원칙을 고려하여 만들어져야 한다. 그리고 단어인지 과정안에는 수많은 하부과정들이 있는 데, 이 하부과정들도 가능한 단원적 설계의 원칙 (principle of modular design) 에 따라 구성되어야 할 것이다.

단어인지 과정에서 활성화되는 정보의 종류가 시간경과에 따라 다른 선행연구들에서 누차 지적된 문제이다 (Swinney, 1979). 따라서, 한글 단어의 인지과정에 관한 모델은 어떤 정보가 언제 활성화되는가를 정확하게 기술할 수 있어야 하며, 특히 한국어의 정서법에 관련된 정보들이 언제 어떻게 사용되는가를 명확히 기술해야 할 것이다.

위 두가지 조건을 만족시키면서, 여러 하위 과정의 동시 진행, 병렬 분산 처리등을 훌륭하게 기술하는 모델로서 가장 먼저 고려하고 싶은 모델은 상호 활성화 (Interactive Activation) 모델이다. 이 모델은 영어 이외에, 일본어의 단어인지 과정과 난독증의 사례들을 설명하는 모델로서 적절한 모델이 될 수 있음이 지적되어 있다 (Yi, 1989). 이 모델이 한국어의 정서법적 제 특징들을 수용할 수 있는 지, 또는 그렇게 하기 위하여 어떻게 수정되어야 하는 지 흥미있는 문제이다.

참고문헌

Marslen-Wilson, W.D. & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, 10, 29-63.

McClelland, J.L. & Rumelhart, D.E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, 88, 375-407.

Potter, M.C. (1984). Rapid serial visual presentation (RSVP): A method for studying language processing. In D.E.Kieras & M.A.Just (Eds.), *New Methods in Reading Comprehension Research*. LEA.

Swinney, D. A. (1989). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, 18, 645-659.

Taft, M. (1979). Lexical access via an orthographic code: The Basic Orthographic Syllabic Structure(BOSS). *Journal of Verbal Learning and Verbal Behavior*, 18, 21-39.

이 광오 일본어 단어인지 과정에 대한 Connectionist model. 발표준비중.