

# 어휘판단 과제 시 보이는 언어현상의 계산주의적 모델 설계 및 구현

박기남<sup>†</sup> · 임희석<sup>††</sup> · 남기춘<sup>†††</sup>

## 요 약

본 논문은 인지신경과학의 연구 방법으로 주로 사용되는 어휘판단과제(LDT: Lexical decision task) 시 보이는 언어현상을 모사할 수 있는 계산주의 모델(computational model)을 제안한다. 제안하는 모델은 LDT 시 언어와 독립적으로 관찰되는 언어현상인 빈도효과, 어휘성효과, 단어유사성효과, 시각적쇠퇴효과, 의미점화효과, 그리고 반복점화효과 등을 모사할 수 있도록 설계되었다. 실험 결과, 제안한 모델은 통계적으로 유의미하게 빈도효과, 어휘성 효과, 단어유사성 효과, 시각적 쇠퇴 효과 그리고 의미점화 효과를 모사할 수 있었으며, LDT 시 인간 피험자와 유사한 양상의 수행 양식을 보였다.

키워드 : 어휘판단과제, 순환신경망, 계산주의 모델

## Design and Implementation of Computational Model Simulating Language Phenomena in Lexical Decision Task

Kinam Park<sup>†</sup> · Heuseok Lim<sup>††</sup> · Kichun Nam<sup>†††</sup>

### ABSTRACT

This paper proposes a computational model which can simulate peculiar language phenomena observed in human lexical decision task. The model is designed to mimic major language phenomena such as frequency effect, lexical status effect, word similarity, and semantic priming effect. The experimental results show that the propose model replicated the major language phenomena and performed similar performance with that of human in LDT.

Keywords : Lexical Decision Task, Recurrent Neural Network, Computational Model

### 1. 서 론

최근 인간의 인지기능의 처리 과정과 기제를 밝히고 이를 이해하고자 하는 인지심리학(cognitive

psychology)과 신경심리학(neuro psychology) 연구 분야에서 계산주의적 모델을 이용한 연구가 시도되고 있다[1, 2, 3, 4]. 이러한 시도들은 계산주의적 인지 신경과학(computational cognitive neuro science)이라는 학문 분야로 정의할 수 있는데, 계산주의적 인지신경과학에서 계산주의적 모델을 사용하는 장점은 다음과 같다. 첫째, 계산주의적 모델의 여러 가지 변수와 환경의 조작을

† 정 회 원: 고려대학교 컴퓨터교육과 박사과정  
†† 종 신 회 원: 한신대학교 컴퓨터정보소프트웨어학과 교수  
††† 고려대학교 심리학과 교수 (교신저자)  
논문접수: 2006년 1월 11일 심사완료: 2006년 2월 8일  
\* 본 논문은 2005년 산업자원부 뇌신경정보학연구사업 지원에 의하여 수행되었음.

통하여 실험 모델의 전체적 검정이 가능하다. 둘째, 계산주의적 모델은 경험적 데이터를 해석할 수 있는 토대를 마련해 준다. 즉 표면적으로 모순된 관계를 보이는 경험적 혹은 실험적 데이터이지만 통합된 계산주의적 모델을 통해서 일관성 있게 설명가능하다. 셋째, 계산주의적 모델을 통해 아직 실험으로 밝혀지지 않은 현상에 대한 예측(prediction)이 가능하다. 예를 들면 유아의 단어 습득에 관한 계산주의적 모델의 연구 사례가 있는데, 실험적으로 밝혀지지 않은 현상을 연결주의 모델(connectionism model)을 통해 제시하였고, 이후 그 현상이 실제 유아를 통해 밝혀진 경우가 있다. 마지막으로 인간에게 직접 행하기 어려운 실험을 계산주의적 모델을 통해 수행할 수 있다. 뇌 손상 환자와 같은 경우, 뇌 손상의 정도 혹은 손상 부위를 통제하여 특정 과제를 수행하게 하는 것은 불가능하거나, 윤리적으로 문제가 되지만, 계산주의적 모델을 통해서 그러한 조건 제시가 가능하다.

본 논문은 한국어 어휘정보처리 과정의 계산주의적 모델을 개발하기 위한 연구의 일부로서 어휘판단과제 시 보이는 언어 현상을 모사할 수 있는 계산주의적 모델을 제안하고 이의 구현에 대하여 논한다. 이를 위하여 한국어 어휘판단과제의 계산주의적 모델의 입출력 방식과 신경망을 이용한 계산주의적 모델을 제안한다. 또한 제안한 신경망 모델을 대용량의 코퍼스에서 추출한 실제 한국어 어휘로 학습시키고 어휘판단과제를 실시하여 계산주의적 모델과 인간의 어휘판단과제 수행과의 상관관계를 보임으로서 본 논문이 제안하는 계산주의적 모델의 타당성을 검증하고자 한다.

## 2. 관련 연구

### 2.1. 어휘판단과제

단어재인이란 일종의 대응(mapping)과정이라 볼 수 있는데, 이러한 관점에서 시각단어재인은 시각적 철자자극(orthography)을 의미(semantic)

로 대응하는 과정, 혹은 시각적 철자자극을 음운(phonology)정보를 통해 의미로 대응하는 과정이라 할 수 있다[2, 3]. 언어를 연구하는 언어 심리학자들은 이러한 대응과정, 즉 인지과정의 정보처리 과정을 밝히고자 여러 가지 패러다임을 사용하였다. 그중에서도 어휘판단과제(lexical decision task)를 이용한 연구가 대표적인 연구이다.

어휘판단과제는 1970년에 Harbert Rubenstein에 의해서 처음 사용되었다[5]. 이 실험은 여러 개의 어휘 즉, 문자열이 피험자에게 차례대로 제시 될 때 피험자는 되도록 빠르고, 정확하게 제시된 문자열이 단어인지 아닌지 판단하게 된다. 일반적으로 차례대로 제시되는 문자열은 반은 단어이고, 나머지 반은 단어가 아닌 비단어 피험자에게 나타는 반응변수로는 반응시간과 정확 반응률이 측정된다.

이 연구방법은 단지 제시 문자열이 문자열의 의미에만 접근해야 하기 때문에 다른 부가적인 처리과정을 요구하지 않는다. 따라서 어휘처리 연구에 적절한 방법이라 볼 수 있다. 어휘판단과제를 이용하여 어휘접속과정을 연구한 많은 연구자들은 시각정보가 의미로 대응되는 언어처리과정과 언어처리과정에 관여하는 여러 요소에 대해 밝혀냈다. 어휘판단과제는 넓게는 단어재인, 좁게는 어휘접속과정에 나타나는 안정된 여러 효과들을 보여주고 있으며, 여러 연구들을 통해 어휘접속과정 연구에 적합한 방법으로 제시되어 왔다[3].

어휘판단과제를 통해 밝혀진 여러 효과들은 대부분 많은 연구자들에 의해서 중복적으로 밝혀져 온 효과들이며, Taft가 정리한 효과[6]들을 통해 한국어 언어 사용에서 나타나는 언어현상들을 기술하면 다음과 같다.

### 2.2. 단독으로 제시된 단어의 인지

#### ■ 빈도효과 (frequency effect)

단어의 인식에 있어서 단어인식의 빠르고, 느림은 단어의 출현 빈도에 영향을 받게 되는데, 이는 단어의 사용빈도가 어휘적 친숙도를 반영하

고, 친숙도는 다시 단어 인식에 영향을 미친다고 할 수 있다. 즉 사용 빈도가 높은 단어일수록 어휘 판단을 하는데 있어서 사용 빈도가 낮은 단어보다 반응시간이 짧게 나타난다. 직관적인 예를 들자면, 한국어 언어 사용에 있어서 사용빈도가 높은 ‘과일’이 사용빈도가 낮은 ‘과실’보다 어휘 판단 반응 시간이 더 빠르다.

### ■ 어휘성효과 (lexical status effect)

어휘성 효과란 단순히 비단어는 실제 단어와는 다르게 처리된다는 것인데, 즉 어휘 판단에 있어서 단어를 단어라고 판단하는 반응시간이 단어가 아닌 비단어를 단어가 아니라고 판단하는 반응시간보다 짧게 나타나는 현상이다. 예를 들어, 단어가 아닌 ‘칠상’을 단어가 아니라고 판단하는 어휘 판단 반응시간보다 ‘책상’을 단어라고 판단하는 어휘 판단 반응 시간이 더 짧게 나타난다.

### ■ 철자조합적합성효과 (non-word legality effect)

철자조합적합성효과는 언어를 한국어에 국한시켜 설명하면, 한국어 언어사용에 있어서 비단어 인지과정에서 나타나는 효과를 말하는데 한국어 문법규칙, 즉 자소조합에 맞지 않는 비단어를 판단하는 반응시간이 한국어 문법규칙에 맞는 비단어를 판단하는 반응시간보다 짧아지는 현상이다. 예를 들어, 한국어 문법규칙의 자소조합에 맞지 않는 비단어 ‘뽕뽕’의 어휘 판단 반응시간이 한국어 문법규칙의 자소조합에 맞는 ‘빌베’보다 더 짧게 나타난다.

### ■ 단어 유사성효과 (word similarity effect)

단어 유사성효과는 언어를 한국어에 국한시켜 설명하면, 한국어 언어사용에 있어서 비단어 인지과정에서 나타나는 효과를 말하는데, 이는 단어와 비단어의 유사성 정도에 따라 어휘 판단 반응 시간이 다르게 나타난다. 즉 단어와 유사한 비단어의 어휘 판단 반응시간이, 단어와 전혀 무관한 철자조합의 비단어 판단 반응시간보다 더 느리게 나타나는 현상이다. 예를 들어 단어 ‘책상’과 유사한 비단어 ‘칠상’의 어휘 판단 반응시간

이 단어와 무관한 ‘칠쌍’의 어휘 판단 반응 시간보다 더 느리게 나타난다.

### ■ 시각적 쇠퇴효과 (visual degradation effect)

시각적으로 식별이 어려운 어휘일 경우 그렇지 않은 경우보다 어휘 판단 반응 시간이 느려지는 현상을 말한다. 예를 들어 인쇄 상태가 좋지 않거나, 글자가 왜곡된 경우 인쇄 상태가 좋거나, 글자의 왜곡이 없는 어휘에 비해 어휘 판단 반응 시간이 더 느리게 나타난다.

## 2.3. 문맥과 단어인지

### ■ 의미점화 효과 (semantic priming effect)

시간적으로 먼저 제시된 단어가 나중에 제시된 단어의 처리에 영향을 주는 현상을 점화효과 (priming effect)라 한다. 먼저 제시된 단어를 점화단어(prime word)라 하고 나중에 제시된 단어를 목표어(target word)라고 하는데, 의미점화효과는 점화단어와 표적단어가 연상관계에 있을 때에 일어나는 문맥적 효과이다. 예를 들어, 표적단어와 의미적으로 연관된 점화단어 ‘학교’가 제시되고 표적단어 ‘학생’을 제시한 조건과, 표적단어와 의미적으로 연관이 없는 점화단어 ‘병원’을 제시하고 표적단어 ‘학생’을 제시한 조건을 비교해보면 동일한 단어인 ‘학생’임에도 불구하고 어휘 판단 반응 시간은 점화단어가 ‘학교’일 때 더 빠르다.

의미점화효과의 근거로서는 활성화의 전파 (spreading activation)가 가장 유력한 것으로 추측되고 있다. 이는 심성어휘집의 단어들이 네트워크인 의미망(semantic network)을 만들고 있으며, 한 단어의 활성화(activation)는 네트워크를 통해서 이웃(neighborhood)의 다른 단어에게 전파되어 간다는 추론에 기초한 것이다.

### ■ 반복점화효과 (repetition priming effect)

반복적으로 제시된 단어는 그렇지 않은 경우보다 어휘 판단 반응 시간이 더 빨라지는 현상이다.

직관적 예를 들어 ‘책상’을 제시하고 일정 경과한 후 다시 ‘책상’의 의미 파악을 하는 어휘판단 반응 시간이 ‘책상’을 제시하여 단어의 의미를 파악하는 어휘판단 반응 보다 더 빠르게 나타난다.

### 3. 어휘판단과제의 계산주의적 모델

본 논문은 순환 신경망 모델을 이용한 한국어 어휘판단과제의 계산주의적 모델과 신경망을 이용한 어휘정보처리를 위한 한국어의 입출력 방법을 제안한다. 제안하는 모델은 2음절 단어 철자(orthography)입력을 통해 단어의 의미(semantic)를 출력하는 구조를 갖고 있으며, 모델은 철자입력정보를 의미출력정보로 대응시키는 학습을 하게 된다. 학습이 이루어진 후 모델은 어휘판단과제를 모사하기 하기 위해 학습된 데이터와 학습되지 않은 데이터의 철자입력정보를 토대로 의미출력정보를 산출하게 된다.

모델에서 학습된 문자열은 단어조건이 되고, 학습되지 않은 문자열은 비단어 조건이 된다. 모델이 산출한 각 조건별 의미 출력값은 의미강조값(semantic stress value)으로 변환하여 각 조건별 의미강조값의 차이를 통해 어휘판단과제에서 나타나는 언어 현상들을 검증하고자 한다.

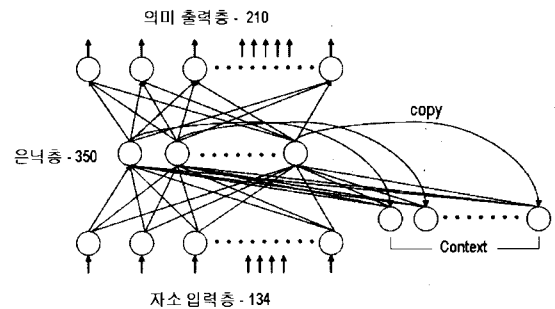
#### 3.1. 계산주의적 모델의 구조

본 논문에서 제안된 모델은 <그림 1>과 같이 3개의 층을 가진 단순한 형태의 단순 순환 신경망 모델(simple recurrent neural network)로 다층 퍼셉트론 모델에 은닉층의 값을 보관하는 문맥층(context)을 추가하고, 회기(feedback) 연결을 추가함으로써 신경망이 동적 특성을 갖도록 한 Elman's Simple Recurrent Network 구조로 설계하였다.

<그림 1>은 제안한 모델의 구조를 보이고 있으며, 자소 입력층은 134개의 단위(unit), 은닉층은 350개의 단위(unit), 의미 출력층은 210개의 단위(unit)로 구성되어있다. 여기서 입력층은 문맥 유닛(context unit)과 입력 유닛(input unit)으로 구성되며, 문맥 유닛은 은닉층의 값을 한 번

의 시간(interval)간격 후에 입력으로 전달하는 역할을 하며 은닉층의 유닛개수와 동일하다. 초기 문맥층은 0값을 갖도록 설정하였다. 모델 구조에서 각 층(layer)은 완전(fully connect)하게 연결되었으며, 은닉층과 출력값은 입력벡터와 연결 가중치벡터(weight vector)를 내적(inner product)하여 계산한다.

각 단위의 출력함수(output function)는 비선형 함수인 시그모이드(sigmoid) 함수를 사용하였다.



<그림 1> 어휘판단과제에 계산주의적 모델

#### 3.2. 입력 설계

입력 데이터 구조에 대한 논의와 문제점은 이미 기존 연결주의 모델에서 많이 논의되어왔다[1, 2, 3]. 하지만 한글의 경우 초성, 중성, 종성의 구분이 영어의 알파벳처럼 단지 위치적 구분이 아닌 언어학적인 구분이기 때문에 자소조합 자체만으로 언어학적 측면을 고려한 입력패턴 설계가 가능하다. 이는 기존 연결주의 모델의 입력 데이터 구조 설계에 대한 문제점들을 해결함은 물론 한글 체계에 대한 특징을 반영해 줄 수 있다. 물론 음절을 표상한 입력 데이터 구조가 실제 언어 환경을 완벽하게 반영하지는 못하지만, 본 논문이 제안한 모델에서 보여주고자 하는 언어현상을 설명하기에는 충분하다고 본다.

컴퓨터에서 한글의 한 음절은 조합형 글자체계에서와 같이 초성, 중성, 종성에 각각 5비트(bit)씩을 할당하고, 최상위 비트를 영어 또는 한글을 표시하는 비트로 사용하여 2바이트(byte)로 표현

할 수도 있다. 하지만 본 논문의 입력설계를 이처럼 표현할 경우 서로 다른 음절이 유사한 비트열을 가질 수 있어 신경망의 입력에서 유사한 음절로 취급 될 수 있는 문제점이 있다. 이러한 이유로 본 논문은 1차원에서 19차원까지는 초성에 해당되는 자음(복자음)을, 20차원에서 40차원까지는 중성에 해당되는 모음(이중모음)을, 41차원에서 67차원까지는 중성에 해당하는 자음(복자음)을 나타내도록 하는 총 67차원의 벡터 형식으로 1 음절을 표현하도록 하였다.

본 논문의 실험을 위하여 사용된 2음절 어휘표현은 총 134차원의 벡터를 사용하였고, <표 1>은 한글의 모든 자모와 이에 대응하는 입력값을 나타낸 것이다. 모든 벡터는 0으로 초기화 되어 있고, 자모를 표상하기 위해서 자모에 해당하는 벡터를 1로 설정하였다. 따라서 각 입력 차원들은 서로 독립적(orthogonal)인 관계를 갖는다.

<표 1> 한국어 자모의 벡터표현 설계

초성(19)	
ㄱ	10000000000000000000
ㅋ	01000000000000000000
ㆁ	00100000000000000000
ㄷ	00010000000000000000
:	:
ㅎ	00000000000000000001
중성(21)	
ㅏ	10000000000000000000
ㅑ	01000000000000000000
ㅓ	00100000000000000000
ㅕ	00010000000000000000
:	:
ㅣ	00000000000000000001
종성(27)	
ㄱ	100000000000000000000000
ㅋ	010000000000000000000000
ㆁ	001000000000000000000000
ㄴ	000100000000000000000000
:	:
ㅇ	000000000000000000000001

<표 1>에서 제시된 입력패턴 설계 방법에 따라 2음절 어휘를 표현을 예로 들면, ‘각단’이라는 단어를 표현하기 위해 첫 번째 음절의 초성19차원 중 ‘ㄱ’차원, 중성21차원 중 ‘ㅏ’차원, 종성27차원 중 ‘ㄱ’차원, 두 번째 음절의 초성19차원 중

‘ㄷ’차원, 중성21차원 중 ‘ㅑ’차원, 종성27차원 중 ‘ㄴ’차원을 1로 설정하여야 하며, 나머지 차원은 0으로 설정해야한다. <표 2>는 ‘각단’이라는 2음절 단어를 입력패턴 설계 방법으로 설계한 예를 보인 것이다.

<표 2> 134개의 독립벡터 표현

각단	ㄱ	10000000000000000000	ㅏ	10000000000000000000
	ㅋ	10000000000000000000000000000000	ㅑ	00010000000000
	ㆁ	000000	ㅓ	100000000000000000000000
		000000000000000000000000	ㅕ	000100000000

### 3.3. 출력 설계

어휘판단과제의 계산주의적 모델의 출력은 입력된 단어의 의미값이다. 단어의 의미는 총 210개의 의미 자질로 구성된 벡터로 표현된다. 출력 설계에 해당하는 의미의 표상은 실제 단어의 의미를 정확하게 고려하여 설계할 수 있지만, 단순화를 위해 실제 의미를 표상하여 설계하지 않고, 다만 입력문자와 의미간의 자의성과 단어의 의미들간의 범주화를 고려하여 만들어졌다.

출력 설계 시 문자와 의미간의 자의성은 단어 자소배열의 유사성이 의미의 유사성을 반영하지 않도록 하였다. 예를 들어 ‘사람’과 ‘사랑’은 철자(자모)간의 유사성은 크지만, 그 의미는 전혀 유사하지 않다. 따라서 철자를 토대로 만든 입력패턴을 랜덤하게 생성시킨 의미패턴에 대응 시키는 것은 입력패턴구조와 의미패턴구조의 관계를 추상적으로 모사했다고 할 수 있다. 또한 의미패턴구조에 대한 범주화 고려는 의미패턴 간의 유사성을 반영하는 것으로 의미간의 관계를 추상적으로 모사 할 수 있다.

<표 3>은 의미적으로 연관이 있는 어휘를 의미패턴구조로 30개씩 의미 범주화 하여 나타내었고, 문자와 의미간 자의성과 의미표상의 범주화를 위해 210차원을 이용하여 7개의 프로토타입(prototype)을 생성시켰고, 각 차원이 1로 활성화 될 확률은  $P_p=0.15$  이다. 이렇게 만들어진 7개의 프로토타입을 기준으로 건본(exemplar)을 29개씩 생성시켜 어휘의 의미패턴을 설계하였다.

<표 3> 의미간의 관계를 표현한 의미표상구조

	1	...	30	31	...	60	61	...	90	91	...	120	121	...	150	151	...	180	181	...	210	
학생	1		10111...11011			000100...00000			000100...00000			000000...00000			000010...00000			000000...00000			000000...00000	
~			:			:			:			:			:			:			:	
의사	30		11011...10101			000000...00010			001000...00000			000010...00000			000000...00000			000000...00000			000000...00000	
사과	31		000000...00000			10111...10101			000000...00000			000000...00000			000000...00000			000000...00000			000000...00000	
~			:			:			:			:			:			:			:	
뽕기	60		000010...00000			10101...10110			000000...00000			000000...00000			000010...00000			000000...00000			000000...00000	
책상	61		000000...00000			000100...00000			10111...01111			000100...00000			000000...00000			000100...00000			000100...00000	
~			:			:			:			:			:			:			:	
의자	90		000000...00000			000000...00010			11101...11101			000000...00010			000000...00000			000000...00010			000000...00010	
과정	91		000000...00000			000000...00000			000000...00000			11011...11101			000000...00000			000100...00000			000100...00000	
~			:			:			:			:			:			:			:	
시간	120		000000...00000			000010...00000			000000...00000			10111...11101			000000...00000			000000...00010			000000...00010	
다리	121		000000...00000			000100...00000			000000...00000			000000...00000			11011...11011			000000...00000			000000...00000	
~			:			:			:			:			:			:			:	
머리	150		000010...00000			000000...00010			000010...00000			000000...00000			11011...11101			000000...00000			000000...00000	
종아	151		000100...00000			000000...00000			000100...00000			000000...00000			000100...00000			11110...01111			000100...00000	
~			:			:			:			:			:			:			:	
유리	180		000000...00010			000000...00000			000000...00010			000000...00000			000000...00010			11101...10111			000000...00010	
느낌	181		000100...00000			000000...00000			000100...00000			000000...00000			000100...00000			000000...00000			10111...11110	
~			:			:			:			:			:			:			:	
사랑	210		000000...00010			000000...00000			000000...00010			000000...00000			000000...00010			000000...00000			11101...10111	

3.4. 계산주의적 모델의 학습

본 논문에서 제안한 모델을 학습시키기 위해 학습 데이터의 입력데이터는 세종 말뭉치(1999, 2000, 20001년)에서 추출한 2음절단어 총 30,315개를 이용하여 만들어졌다. 각각의 단어는 세종 말뭉치 내에서 자체 빈도를 고려하여 추출하였고, 출현빈도 5회 이하의 저빈도 70개, 100회이상, 300회 이하의 중빈도 70개, 1000회 이상의 고빈도 70개로 분리하여, 210개를 <표 4>에서 제안한 입력패턴설계 방법을 이용하여 학습데이터를 구축하였다.

출력값은 입력된 단어의 의미표현이기 때문에 210개의 입력데이터와 같은 수를 의미패턴설계구조를 바탕으로 만들어졌고, 입력데이터와 출력데이터는 랜덤하게 짝지어졌다. 이는 단어들 간의 철자조합 유사성과 의미간의 관계가 존재하지 않음을 나타내 주기 위한 것이다.

본 논문은 제안한 모델을 학습시키기 위해 오류 역전과 학습 알고리즘을 사용하였다. 가중치 변경은 오직 전방향의 연결에 대해서만 이루어지기 때문에, 목표 출력값(출력데이터)과 실제 출력 출력의 차이를 계산하여 오차 값을 전달하는 오류 역전과 알고리즘을 그대로 사용할 수 있다. 다만, 학습과정에서 출력되는 오류 값은 일반적인

신경망의 입력값과 목표값의 차이를 평균 제공한 RMS(root mean square)를 사용하지 않고, <식 1>에 의한 교차엔트로피(cross entropy)값을 사용하였다. <식 1>에서 E는 오류값(error value),  $o_d$ 는 노드d에 대한 출력값,  $t_d$ 는 d에 대한 목표 값을 나타내며 D와 d는 각각 전체 학습데이터와 개별 학습 데이터를 의미한다.

$$E = - \sum_{d \in D} \sum_i (t_{d_i} \log o_{d_i} + (1 - t_{d_i}) \log (1 - o_{d_i})) \quad \text{<식 1>}$$

제안한 모델은 1,000번의 시행 학습에 학습율(learning rate) 0.1, 가중치 변경 알고리즘은 Doug's 모멘텀(momentum) 0.9를 사용하였다. 출력력은 매 10회 학습에 학습 데이터에 대한 오류값을 출력하도록 설정하였다[7].

3.5. 실험데이터

본 논문은 제안하는 모델의 평가를(evaluation) 위하여 단어와 비단어를 사용하였다. 실험에 사용된 단어와 비단어 모두 실제 언어생활에 쓰이는

단어이지만, 본 논문에서 제안한 모델이 인간의 언어처리과정을 모사한 것이므로 모델의 학습에 참여한 단어를 ‘단어’, 학습에 참여하지 않은 단어를 ‘비단어’라고 정의한다.

모델의 평가를 위한 실험데이터는 <표 4>를 바탕으로 단어와 비단어, 단어와의 유사성, 의미적 연관성, 시각적 인식률 그리고 출현빈도를 고려하여 설계하였다.

실험 데이터 조건들은 목표값 없이 설계되었고, 학습을 마친 모델에서 1회 계산되어 모델의 출력값을 이용하여 분석하였다.

<표 4> 실험 데이터 설계과정

LDT 평가	실험데이터 설계	
빈도효과	단어 210개, 고빈도 단어 70개, 저빈도 단어 70개	학습에 참여한 단어
어휘성 효과	단어 70개 비단어 70개	학습에 참여한 단어 학습에 참여하지 않은 비단어
단어 유사성효과	단어와 유사한 자소조합으로 이루어진 비단어 70개	예, “ 책상 ”
시각적최취효과	단어와 유사하지 않은 조소조합으로 이루어진 비단어 70개	예, “ 칠상 ”
의미접화 효과	단어 70개 시각적으로 인식이 어려운 단어(partial) 70개	예, “ 학생 ” 예, “ 학생 ”
의미접화 효과	표적(target)단어와 의미적으로 연관된 점화(prime)단어와 표적단어 쌍 70개	예, “ 학생 & 책상 ”
의미접화 효과	표적(target)단어와 의미적으로 연관되지 않은 점화(prime)단어와 표적단어 쌍 70개	예, “ 병원 & 책상 ”

#### 4. 실험결과

본 논문에서 제안한 모델의 출력값은 의미강조값(semantic stress)값을 이용하여 분석되었다. 의미강조값은 엔트로피값의 일종으로 오류값을 계산하는 방법 중 하나이다. 오류값이 크다는 것은 의미강조값이 낮다는 것이고, 어휘판단과제 시 어휘판단 반응시간이 길다는 것이며 오류값이 작다는 것은 의미강조값이 크다는 것이고, 어휘판단 반응시간이 짧다는 의미이다.

의미강조값의 계산식은 <식 2>과 같고, 의미층에서의 출력값이 0에서 1사이의 값 중 0.5에 가까울수록 0이 되고, 0이나 1에 가까울수록 1이 된다. 이는 본 모델에서 입력설계구조와 출력설계구조

가 0혹은 1로 설계되었다는 점과 실험데이터가 목표값이 없어 출력값과 목표값의 차이를 평균 제공한 값을 이용할 수 없다는 점에서 모델평가 함수로서 적절하다.

$$S_j = s_j \log_2(s_j) + (1 - s_j) \log(1 - s_j) - \log_2(0.5) \quad <식 2>$$

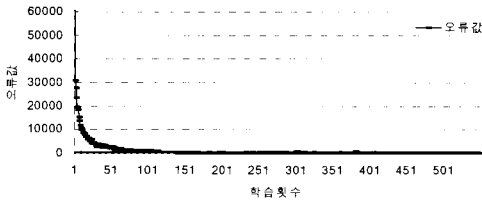
$S_j$ : 의미강조값  $s_j$ : 의미층에서의 출력값

1,000회의 학습시행 및 10,000회의 업데이트(update)를 마친 모델은 교차엔트로피함수에 의해 계산된 총 오류수가 0.15로 한 입출력 당 평균 0.0005의 매우 작은 값이었다. 표면적으로는 성공적인 학습이 이루어진 것 같지만, 신경망 학습의 특성상 과 학습(over fitting)을 고려해야 하기 때문에 학습데이터의 단어에 대한 상대빈도와 학습 횟수에 따른 의미강조값을 <식 3>를 이용하여 피어슨 단순 상관계수(Pearson’s simple correlation coefficient)를 계산한 후 가장 상관이 높은 지점에서의 학습 결과를 이용하여 신경망을 실험에 사용하였다.

$$r = \frac{\sum(f_i - \bar{f})(o_i - \bar{o})}{\sqrt{\sum(f_i - \bar{f})^2 \sum(o_i - \bar{o})^2}} \quad <식 3>$$

$r$ : 피어슨(Pearson’s)상관계수,  $f_i$ : 학습 단어의 상대빈도,  $o_i$ : 학습횟수에 따른 의미강조값

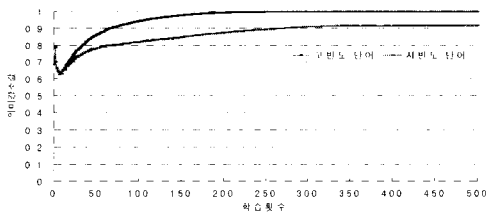
<식 3>를 이용하여 학습 단어의 상대빈도와 학습 횟수에 따른 의미강조값 간의 상관계수를 계산한 결과, 500번째 학습시행 때 상관이 가장 높았으며 이때의 피어슨 단순 상관계수는 0.833 ( $p < 0.01$ )이었고, 모델의 총 오류수는 0.234이었으며, 한 입력 당 평균 0.00078이었다. 또한 <그림 2>은 학습시행 횟수에 따른 오류값의 변화를 보여주는 데 학습 횟수가 증가할수록 오류값이 감소함을 보임으로써 학습이 잘 이루어졌음을 나타낸다.



<그림 2> 학습횟수에 따른 오류값의 변화

#### 4.1. 빈도효과

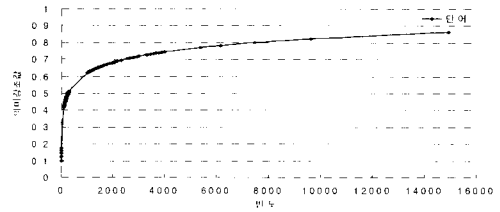
본 논문에서 제안한 모델에 의해 어휘판단과제의 빈도효과를 모사하기 위해서는 먼저 학습단어들의 빈도와 의미강조값들 사이에 유의미한 상관관계가 나타나야한다. 즉, 학습단어에서 제시한 빈도 증가에 따라 의미강조값이 증가하는 정적 상관관계가 보여 진다면, 모델은 빈도효과를 모사했다고 할 수 있다. 다음으로 저빈도 단어들과 고빈도 단어들 사이에 의미강조값 평균이 유의미한 차이를 보여야 한다. 분석결과 학습된 단어의 빈도와 의미강조값의 정적인 상관이  $0.833 (p < 0.01)$ 로 유의미하게 나타났고, 학습된 단어의 의미강조값 평균은 <그림 3>와 같이 학습 횟수에 따른 차이를 보였으며, 각각 고빈도 단어  $0.960$ , 저빈도 단어  $0.865$ 로 유의미한 차이를 보였다 ( $t[998] = 20.721, p < 0.05$ ).



<그림 3> 학습 횟수에 따른 고빈도, 저빈도 단어의 의미강조값 평균

<그림 4>는 단어 빈도와 의미강조값과의 상관관계를 보이고 있는 것으로 단어의 빈도가 높아짐에 따라 의미강조값이 높아짐을 볼 수 있다. 높은 의미강조값은 짧은 반응 시간을 의미하는 것으로 어휘판단과제 시 인간이 보이는 빈도 효

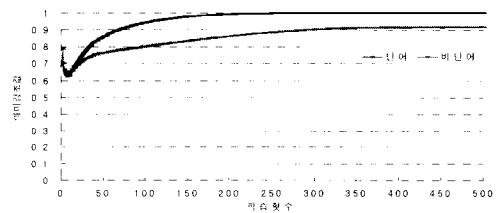
과를 보이고 있는 것으로 해석할 수 있다.



<그림 4> 단어 빈도와 의미강조값의 상관관계

#### 4.2. 어휘성 효과

어휘성 효과에 대한 분석 결과, 데이터의 의미강조 값 평균은 <그림 5>과 같이 학습 회수가 증가함에 따라 단어와 비단어간에 차이를 나타냈으며, 각각 단어  $0.960$ , 비단어  $0.858$ 로 유의미하게 차이가 났다 ( $t[998] = 21.902, p < 0.05$ ). 따라서 학습된 모델은 어휘성 효과를 보였다.



<그림 5> 학습횟수에 따른 단어, 비단어의 의미강조값 평균

#### 4.3. 단어 유사성효과

모델의 단어 유사성효과 모사 여부는 단어와 유사한 철자조합의 비단어와 단어와 전혀 다른 구조의 철자조합 간의 어휘판단 반응 시간을 차이를 살펴봄으로써 판단할 수 있다. 모델의 평가는 학습단어와 유사한 철자조합 구조의 비단어 70개와 단어와 전혀 다른 철자조합 구조의 비단어 70개의 단어조건을 사용하였다.

단어와 비단어간 철자조합 유사도는 <식 4>와 같이 단어를 나타내는 입력 벡터간의 코사인 유사도를 이용하여, 학습단어에서 무작위로 뽑은

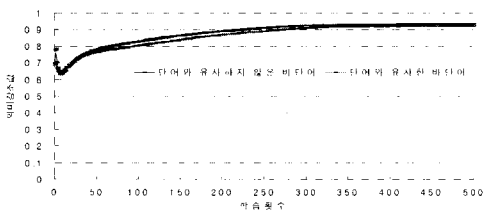


70개의 단어와 유사한 철자조합 비단어의 유사도가 단어와 유사하지 않은 철자조합 비단어의 유사도의 2배 이상이 되도록 설정하였다. 이때 각각 코사인 유사도(cosine similarity)는 단어와 유사한 철자조합의 비단어 0.104, 단어와 유사하지 않은 철자조합 비단어 0.041을 갖고 있었다 ( $t[71.316] = 5.354, p < 0.05$ ).

$$sim(x, y) = \frac{\sum_{i=1}^t x_i y_i}{\sqrt{\sum_{i=1}^t x_i^2 \sum_{i=1}^t y_i^2}} \quad \text{〈식 4〉}$$

$sim(x, y)$ : 코사인 유사도,  $x$ : 단어,  $y$ : 비단어,  $t$ : 차원

분석결과 <그림 6>와 같이 학습 횟수에 따른 의미강조값의 차이가 단어와 유사한 철자조합 구조의 비단어 경우 의미강조값 평균이 0.862이고, 단어와 전혀 다른 철자조합 구조의 비단어 경우 의미강조값 평균이 0.878로 유의미한 차이를 보였다 ( $t[998] = 3.768, p < 0.05$ ). 모델의 의미강조값이 단어와 유사한 철자조합구조의 단어보다 단어와 전혀 다른 철자조합 구조의 비단어 일때 더 크게 나타나 모델은 단어 유사성효과를 잘 모사하고 있다.

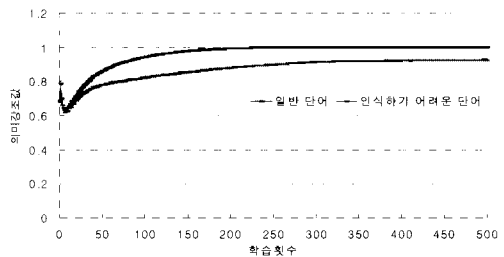


<그림 6> 학습횟수에 따른 철자구조가 단어와 유사한 단어와 단어와 유사하지 않은 단어의 의미강조값 평균

#### 4.4. 시각적 쇠퇴효과

모델이 시각적 쇠퇴효과를 보이기 위해서는 시각적으로 식별이 어려운 단어일 경우 단어를 판단하는 어휘판단반응시간이 정상 단어를 판단하는 경우보다 어휘판단반응 시간이 길게 나타나야 한

다. 모델을 평가하기 위해서 학습에 참여한 단어 70개를 선택하여, 시각 쇠퇴 효과를 표상하기 위해 입력벡터를 한 단어 당 30%(40벡터)를 1일 아닌 0.5로 설정하였다. 분석결과 <그림 7>과 같이 학습 횟수에 따라 의미강조값의 평균이 변화에 차이가 있었으며, 정상 단어일 경우 의미강조값 평균이 0.960이고, 비정상 단어(partial word) 경우 의미강조값 평균이 0.871로 유의미한 차이를 보였다 ( $t[998] = 19.182, p < 0.05$ ). 따라서 본 모델이 어휘판단과제의 시각적 쇠퇴효과를 잘 모사하였다.



<그림 7> 학습횟수에 따른 시각적 정상인식 단어와 비정상 인식 단어의 의미강조값 평균

#### 4.5. 의미점화 효과

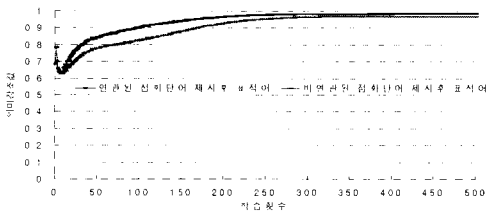
본 논문의 전문에서 언급했듯이 의미점화효과가 활성화 전과 이론에 바탕을 두고 있기 때문에 점화단어가 표적단어에 영향을 미치게 된다. 즉 표적단어와 의미적으로 연관된 점화어가 제시되었을 때 의미적으로 연관성이 없는 점화화어가 제시되었을 때 보다 빠른 어휘판단 반응 시간을 보이면, 모델이 의미점화 효과를 잘 모사했다고 할 수 있다.

분석결과 의미적으로 연관된 점화단어 제시 후 표적어의 의미강조값 평균이 0.939이고, 의미적으로 연관되지 않은 점화단어 제시 후 표적어의 의미강조값 평균이 0.903로 유의미한 차이를 보였다 ( $t[974.472] = 7.048, p < 0.05$ ). 학습 횟수에 따른 의미강조값의 변화량은 <그림 8>와 같이 차이를 보였다. 따라서 모델이 의미점화 효과를 잘 모사 했고, 심성어휘집의 단어들이 의미망을 구성하고, 단어의 활성화가 이웃의 비슷한 의

미의 단어에게 전파 된다는 이론을 뒷받침할 수 있다고 볼 수 있다.

### 5. 결론

본 논문은 한국어 언어처리 과정을 연구하기 위한 어휘판단과제 시 보이는 언어현상을 모사할 수 있는 계산주의적 모델을 제안하고, 제안된 모델의 타당성을 검증하기 위하여 실험을 실시하였다.



<그림 8> 학습횟수에 따른 의미적으로 연관된 단어가 접두어로 사용된 경우와 의미적으로 무관한 접두어가 사용된 후 표적단어의 의미강조값 평균

실험결과 학습단어의 의미강조값과 모델 출력의 의미강조값들 사이의 유의미한 상관관계와 고빈도 단어와 저빈도 단어간의 의미강조값 차이로 어휘판단과제에서 보이는 빈도효과를 설명하였고, 단어와 비단어 간의 의미강조값의 차이를 통해 어휘성 효과를 모사하였으며, 비단어간 철자조합 구조에 따른 차이로 단어유사성 효과를 설명하였다. 또, 실험데이터의 입력 값을 조작함으로써 어휘처리의 시각최퇴효과를 보였다.

마지막으로 접두어단어와 표적어간의 활성화 정도를 통해 활성화 전파(spreading activation)이론을 모사함으로써 어휘판단과제의 의미점화효과를 설명하였다. 이는 모두 의미강조값을 반응시간의 준거로 가정하고 모델이 이러한 준거를 토대로 의사결정 할 것이라는 전제를 바탕으로 하고 있다.

본 논문은 인간의 시각단어재인 과정 중 어휘판단과제 시 보이는 언어현상을 정보처리론적 계산주의적 모델을 통해 모사함으로써 한국어 언어 정보처리의 인지신경학적 연구가 미미한 상태에서 새로운 연구 결과를 통해 한국어 언어 정보처리의 인지신경학적 연구의 근간을 마련하였다. 하

지만, 본 연구가 인간의 언어처리 과정의 수많은 기제 중 한부분의 연구임을 감안할 때 추후 연구를 통해 더 많은 연구 기제를 해결함과 동시에 더 많은 설득력을 갖게 해볼 수 있을 것이다.

특히, 신경망은 인간의 뇌처럼 입력과 출력 신경세포 사이의 신경연결을 논리에 의해 모형화할 수 있기 때문에 인간 신경을 논리적 진술로, 신경의 흥분 또는 억제를 표현할 수 있다. 따라서 신경망을 이용한 계산주의적 모델을 통하여 인간 뇌의 여러 기제 중 뇌 손상 환자를 모델링하고, 전부 또는 일부를 손상시켜 좁게는 뇌손상 환자를 모사하여 뇌 손상 이후의 현상을 예측할 수 있고, 더 나아가 치료에 도움이 될 수 있는 방법론을 제시 할 수 있을 것이다. 또 한 언어학적 연구뿐만 아니라 연구 도메인을 다르게 하여 실질적 응용분야 적용함으로써 연구의 더 큰 의의를 찾을 수 있을 것이다.

### 참 고 문 헌

- [1] Plaut, D. C., "Relearning after damage in connectionist networks", Toward a theory of rehabilitation. *Brain and Language*, 52, 25-82, 1996.
- [2] Plaut, D. C., "Insights from distributed models of word reading and lexical decision", *Structure and Function in the lexical system*, Ed., *Language and Cognitive Processing*, pp.765-805, 1997.
- [3] Rubenstein, H., Lewis, S. S., & Rubenstein, M. A., "Evidence for phonemic recoding in visual word recognition", *Journal of Verbal Learning and Verbal Behavior*, 10, pp.645-657, 1971.
- [4] Ellis, R., & Humphreys, G., *connectionist psychology: A text with readings*, Hove, East Sussex, Psychology Press. 1999.
- [5] Rubenstein, H., Lewis, S. S., & Rubenstein, M. A., "Evidence for phonemic recoding in visual word recognition", *Journal of Verbal Learning and Verbal Behavior*, 10, pp.645-657, 1971.
- [6] Taft, M. *Reading and the mental lexicon: School of Psychology University of New South Wales Australia*, 1991.
- [7] Rumelhart, D. E. & McClelland, J. L., *Parallel*

distributed processing-explorations in the microstructure of cognition: The MIT press, 1, 1986.

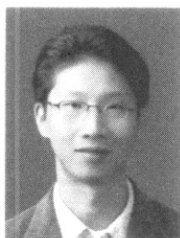
### 박 기 남



2004. 2 : 천안대학교  
컴퓨터학과 (학사)  
2006. 2 : 한신대학교  
컴퓨터정보학과 (석사)  
2006 ~ 현재 : 고려대학교  
컴퓨터교육학과 박사과정

관심분야: 데이터 마이닝, 인지신경계산학  
E-Mail : [spknn@korea.ac.kr](mailto:spknn@korea.ac.kr)

### 임 희 석



1992. 2 : 고려대학교  
컴퓨터학과(학사)  
1994. 2 : 고려대학교  
컴퓨터학과(석사)  
1997. 9 : 고려대학교  
컴퓨터학과(박사)

1997. 9 ~ 1999. 2 : 삼성종합기술원 전문연구원  
1999. 3 ~ 2004. 2 : 천안대학교 정보통신학부 교수  
2004. 3 ~ 현재 : 한신대학교 컴퓨터정보소프트웨어학부 교수  
관심분야 : 자연어처리, 인공지능, 인지신경계산학  
E-Mail : [limhs@infocom.chonan.ac.kr](mailto:limhs@infocom.chonan.ac.kr)

### 남 기 춘



1985. 2 : 고려대학교  
심리학과 (학사)  
1988. 2 : 고려대학교  
심리학과 (석사)  
1995 : The Univ.  
of Texas at Austin Ph.D.  
1998. 3 ~ 현재 : 고려대학교

심리학과 교수  
관심분야 : 인지신경학, 음성처리, 언어심리학  
E-Mail : [kichun@korea.ac.kr](mailto:kichun@korea.ac.kr)