

직관 실험 및 코퍼스를 바탕으로 한 의미 중의성 해소 계산 모형 연구*

A Study on the Computational Model of Word Sense Disambiguation,
based on Corpora and Experiments on Native Speaker's Intuition

김 동 성**
(Dong-Sung Kim)

최 재 웅***
(Jae-Woong Choe)

요 약 본 논문은 의미 중의성 해소에 대한 화자의 직관의 계산 모형에 대한 연구로, Harris (1964)의 ‘분포가설’에 근거하여 핵심어와 공기하는 어휘들에 대한 분포적 정규성을 포착하는 언어 직관의 계산 모형을 제안한다. 이를 위해 분포적 정규성에 대한 화자의 처리 계산 모형을 파악하기 위하여 심리언어학적 실험을 실시하고 그 결과를 분석한다. 계산 모형으로는 논리 모형, 확률 모형, 그리고 확률 추론 모형의 세 가지 모형이 설정되었다. 실험은 두 가지로 구성되었다. 첫 번째는 100만 어절 코퍼스에서 추출된 문장을 화자 직관으로 의미를 식별하는 실험이었다. 이 실험에서는 응답간 일치도가 98%로 나왔다. 두 번째 실험은, 제한된 환경에서 실험자의 반응을 관찰하기 위한 것으로, 분열문이라는 환경을 통해 핵심어와 공기어 사이의 의미 관계를 살펴보았다. 또한 100만 어절 코퍼스에서 관찰된 수치와 실험에서 관찰된 관찰치사이의 상관성을 피어슨의 상관계수로 측정하였다. 그러한 측정 결과 실제 코퍼스에서 관찰되는 현상은 논리 모형과 상관성이 있었고, 제한된 환경에서 실시한 결과는 확률 모형과 상관성이 있었다. 이 실험결과는 논리 모형이 우선적으로 의미 분류에 관여하나, 만약 논리 모형이 적용되지 않을 경우 확률 모형이 관여함을 보여 준다. 아울러 의미 결정 모형의 관점에서는 논리 모형이 정확하게 직관모형을 예측할 수 있었고, 확률 추론 모형도 직관모형을 근사치에 가깝게 예측할 수 있었다.

주제어 의미 중의성 해소, 언어 직관 실험, 부울 논리 모형, 최대우도추정치 모형, 베이지안 확률 모형, 코퍼스

* 이 논문은 2005년 정부(교육인적자원부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구임 (KRF-2005-041-A00275). 또 논문의 일부는 PACLIC 20 (2006.11.1-3, 중국 우한)에서 발표된 바 있음.

** 고려대 언어학과, E-mail: dsk202@korea.ac.kr

*** 교신저자: 최재웅, 고려대학교 언어학과, E-mail: jchoe@korea.ac.kr

Abstract According to Harris(1966) distributional hypothesis, understanding the meaning of a word is thought to be dependent on its context. Under this hypothesis about human language ability, this paper proposes a computational model for native speaker's language processing mechanism concerning word sense disambiguation, based on two sets of experiments. Among the three computational models discussed in this paper, namely, the logic model, the probabilistic model, and the probabilistic inference model, the experiment shows that the logic model is first applied for semantic disambiguation of the key word. Next, if the logic model fails to apply, then the probabilistic model becomes most relevant. The three models were also compared with the test results in terms of Pearson correlation coefficient value. It turns out that the logic model best explains the human decision behaviour on the ambiguous words, and the probabilistic inference model comes next. The experiment consists of two parts; one involves 50 sentences extracted from 1 million graphic-word corpus, and the result shows the agreement rate among native speakers is at 98% in terms of word sense disambiguation. The other part of the experiment, which was designed to exclude the logic model effect, is composed of 50 cleft sentences.

Keywords Word Sense Disambiguation, Experiment, Native Speaker's Intuition, Logic Model, Probabilistic Model, Probabilistic Inference Model, Corpora

1. 서 론

Harris(1964)의 ‘분포 가설(distributional hypothesis)’에 따르면 한 단어의 의미적 특성은 그 단어의 분포적 특성에 기인하고, 그러한 분포적 특성은 주변 단어를 통해서 파악될 수 있다. 본 연구에서는 중의성 해소에 작용하는 분포적 특성 파악의 일환으로 화자 직관에 바탕을 둔 의미 중의성 해소 계산모형을 제시한다.

의미 중의성 해소는 자연어 처리뿐만이 아닌 인공지능 전반에 걸쳐서 기초가 되는 부분이다. 대부분의 의미 중의성 해소 연구는 처리 계산의 효율성이나 의미 구분에 대한 실제 화자들간의 불일치를 중심으로 연구되어 왔다(Ide and Vernois 1998). 그중에서 언어 직관(intuition)을 반영한 계산 모형(computational model)에 대한 연구는 매우 제한적인 부분에서만 이루어졌다(Lapata and Lascarides 2003; Lapata and Brew 2004). 특히 언어 직관을 반영한 일

부 연구에서조차, 주로 확률적인 근거에서만 연구하였을 뿐, 다양한 계산모형의 적합성에 대한 연구는 별로 발견되지 않는다. 즉 언어적 직관이 어떻게 작용하는지, 그리고 그러한 언어 직관을 설명하는 데는 어떠한 계산모형이 적합할 것인지에 대한 보다 체계적인 탐구가 필요하다고 본다.

본 연구는 화자 직관 실험을 바탕으로 한다. 화자 직관 실험을 통해서 측정이 되는 것은 화자 집단의 직관에 따른 선택의 분포와 의미 결정이다. 예를 들어, 중의적 단어 A의 어의 α , β 에 대해서 전체 집단이 8:2로 α , β 를 선택했다면 이는 화자들의 직관 분포이다. 또한 8:2로 우세한 어의 α 가 선택이 되었다면, 이것은 화자 집단의 의미 결정으로, 이러한 의미 결정은 앞서의 직관 분포와 같지 않다. 본 연구에서는 화자 집단의 직관 분포 및 화자 집단의 의미 결정을 기존의 의미 중의성 해소 계산 모형과 비교할 것이다. 그러한 계산 모형이 심리적으로 실재하느냐 하는 문제

는 물론 별도의 연구가 필요하겠으나, 적어도 화자들의 직관을 잘 반영하는 계산모형의 가설에 대한 연구는 나름대로 의의가 있다고 본다.

본 연구에서 비교 검토한 계산 모형은 논리 모형, 확률 모형, 확률 추론 모형이다. 이 모형들은 자연언어 처리 분야의 의미 중의성 해소에서 알고리즘으로 활용되거나, 의미 중의성 해소의 심리실험에 모형으로 제시된 바가 있다. 각각의 모형은 여러 가지 방식으로 의미 중의성 해소에 활용될 수 있다. 모형별로 가능한 여러 방식들 중에서, 본 연구에서는 논리 모형에서 부울 검색(Boolean search), 확률 모형에서 최대우도추정치(maximum likelihood estimation), 확률 추론 모형에서 베이지안 추론(Bayesian inference)을 각각 실험 결과와 비교해 보고자 한다. 이 세 가지 방식이 코퍼스에서 발견되는 측정치와 직관 실험결과가 피어슨 상관관계로 분석이 될 것이다. 이러한 상관관계는 세 가지 모형별로 각기 측정되었고, 그 측정 결과 사이의 상호 유사성과 차이점이 실험 결과와 비교되었다.

계산 모형 확립과 화자 직관 실험을 위해 기존의 코퍼스를 활용하였다. 활용된 코퍼스는 세종계획 2단계 연구 결과 중 한국어 학습용 어휘 선정을 위한 기초 조사에 활용된 ‘현대 국어 사용 빈도 조사’ 코퍼스 100만 어절(이하 [100만코퍼스])이다. 개략적인 절차는, 우선 코퍼스에서 기준에 따라 관련 문장을 추출하였고, 추출된 문장들을 실험참가자들에게 제시하여 주어진 시간 내에 문맥에 적합한 어의를 선택하도록 하였다.

실험은 크게 실제 코퍼스에서 추출된 문장

1) 실험에 사용된 코퍼스는 의미구분이 되어 있는 코퍼스이다.

을 통한 실험과 하나의 중심어와 공기어휘들과 관계를 측정하는 실험, 두 가지로 구성되었다. 첫 번째 실험은 실제 문맥에서의 실험(이하 [실제문맥실험])이고, 두 번째 실험은 그 문맥을 더 제한한 실험(이하 [제한문맥실험])이다. 첫 번째 실험 중 직관 설문 조사에서 실험참가자들간의 직관 일치도가 98%인 것으로 나타났다. 이러한 결과를 설명하는 데는 부울 논리 모형이 매우 적합하다. 그런데, 논리 모형이 적용될 수 없는 경우도 하나가 관찰되었고, 이 경우에 확률 모형이 가장 잘 부합하는 것으로 나타났다. 그러나 단 하나의 사례만 관찰되었다는 것이 이러한 관찰을 일반화하기에는 부족하므로, 그러한 사례의 보편성을 확인하기 위하여 보다 체계적인 조사의 필요성이 제기 되었다. 따라서, 두 번째 실험인 [제한문맥실험]은 논리 모형이 적용될 수 없는 가상적 문맥 환경에서 실험참가자가 어떤 행동양식을 보이는지를 파악하기 위한 것이었다. 그 실험의 결과는 앞의 관찰에서와 같이 확률 모형이 가장 적합한 것으로 밝혀졌다. 따라서 의미 중의성 해소관련 직관의 적용 방식으로, 논리 모형이 우선적으로 적용되고, 논리 모형이 적용되지 않는 경우에 확률 모형이 적용된다는 가설을 세울 수 있었다.

실험을 통해 탐구하고자 한 또 다른 논제는 계산 모형들이 의미 중의성 해소의 결정 모형과 얼마만큼 일치하는가라는 점이다. 즉, 모형이 예측한 것과 실제 실험결과가 얼마만큼 일치하는가라는 문제다. 본 연구에서의 실험 결과, [실제문맥실험]에서 실험참가자들의 직관은 실제 코퍼스에서의 의미 구분과 정확히 일치하였다. [실제문맥실험]에서 밝혀진 것은 부울 검색과 베이지안 추론이 실험참가자들의 결정 모형과 매우 유사하였고, [제한문맥실험]

에서는 베이지안 추론이 실험참가들의 결정 모형과 매우 유사하였다.

본 논문의 구성은 다음과 같다. 2절에서는 의미 분류와 언어학적인 모형이 소개되고, 3절에서는 앞에서 제시한 세 가지의 계산 모형이 제시된다. 4절은 본 연구에서의 실험 대상 및 구성, 절차 등을 자세하게 소개한다. 5절은 실험에 대한 분석과 결과 논의를 담고 있다.

2. 의미 중의성 해소에 대한 기존 연구

2.1 심리학적 연구

‘분포가설’에 따라 핵심어의 의미적 특성은 공기어휘들의 분포적 정규성(distributional regularities)과 연관되어 있다고 본다면, 이러한 특성의 분석을 통해 의미 중의성을 해소할 수 있을 것이다. 어휘 ‘배’는 ‘사람의 신체 기관의 배,’ ‘물위에서 타는 배,’ ‘두 배, 세 배의 배,’ 등 여러 의미를 지닐 수 있다. 그 중에서 어떤 의미로 쓰이는지는 주변 어휘들을 분석함으로써 가능하다. 아래 (1)의 문장에서 ‘배’의 의미는 {케이턴호수, 들어서다}등과 같은 공기어휘에 좌우된다. (부록 [실제문맥실험] 문항]의 19번)

(1) 케이턴호수로 [배]가 들어섰다

화자 직관의 계산 모형은 화자들이 분포적 특성에 기반을 둔 분포적 정규성을 연산하는 것으로 본 연구에서는 가정한다. 즉, 화자 직관의 계산 모형은 이러한 분포적 정규성을 계산하는 인지적 능력을 말하게 된다.

화자의 직관에 대한 기존의 대표적인 연구

에서는 주로 화자간 직관의 불일치 문제를 다루었다. Ahlswede (1995)는 의미 분류에 참여한 실험자간 직관이 서로 불일치할 확률이 60~90%이상으로 넓게 분포하는 것으로 관찰하였다. Vernois (1998)의 연구는 화자직관에 의한 의미분류시 불일치가 발생할 경우에 어떻게 의미를 결정할 것인가에 것이다. Vernois (1998)는 이 연구에서 화자간 의미 분류에 차이가 생길 경우, 화자들의 의미 분류를 카파 통계 (Kappa statistics)를 활용해서 일치 여부를 검증할 것을 주장하였다.²⁾

Lapata and Lascarides (2003)과 Lapata and Brew (2004)는 언어 직관의 선택 모형이 확률적인 계산 모형과 유사함을 입증하였다. 그러나 더 넓은 의미에서의 언어 직관은 ‘분포가설’에 의거하여 주변어를 분석하는 능력으로 분석될 수 있으며, 이를 여러 모형으로 확장해서 비교할 필요성이 있다.

2.2 자연언어 처리 연구

자연언어 처리 분야에서 의미 중의성 해소는 핵심적인 연구분야로 여러 각도에서 광범위하게 연구되고 있다(Manning and Schütze 2000, Ide and Vernois 1998). 이러한 의미 중의성 해소 연구는 기계학습 방식에서 의미 구분된 데이터를 활용하는지 여부에 따라 감독 기반과 비감독기반으로 나뉘게 된다. 본 연구에서 사용된 베이지안 추론 방식은 의미구분이 된 코퍼스를 활용하는 방식으로 감독기반 학습을 활용한다.

본 연구에서 활용된 논리 모형은 부울 대수 연산을 활용한 부울 논리 연산으로 정보 검색

2) 카파 통계는 본 연구에서도 활용되는 바, 5.1절에서 다시 논의될 것이다.

기법에서 사용되는 부울 검색 방식과 유사하다. 이 방식은 해당 정보가 정확히 일치하는 정보를 검색하게 된다(Jackson and Moulinier 2002, Mihalcea and Moldovan 1999).³⁾ 예를 들어, ‘수상 운송기관의 배’의 의미를 지니는 어의의 경우에 ‘타다’와 ‘물’이 공기어휘로서 많이 등장한다. 따라서 중의적 어의가 포함된 문맥에서 ‘타다’와 ‘물’이 포함된 문맥을 찾아내는 것이다. 이 방식에서 가정되는 것은 화자 직관에서 해당 어의와 공기하는 공기어휘의 집합이 심리적으로 실재할 수 있다는 것이다. 예를 들어서 ‘수상 운송기관의 배’라는 어의와 관련하여 {타다, 물}과 같은 어휘집합이 심리적으로 실재한다고 본다. 이러한 어휘군들을 이용해서 의미 중의성 해소를 하는 것은 논리 연산을 통해서 해당 문맥에서 관련 어휘군이 포함되어 있는 경우를 찾아내는 것과 같다. 즉, {타다, 물}과 같은 어휘가 문맥에 존재할 때, ‘수상 운송기관의 배’라는 어의로 분류된다.

그런데, 이러한 어휘들을 활용하는 것은 심리실험에서 문제가 될 수 있다. 만약 해당 어의와 너무 분명하게 연관성이 있는 어휘를 문맥실험에 사용하면 실험의 의미가 크게 감소한다. 즉, ‘수상 운송기관의 배’는 {타다, 물}과 같은 어휘들과의 상관성이 너무 분명해서 실험의 결과가 일방적으로 나오게 될 것이고, 심리실험의 의미를 상실하게 된다. 따라서 실험에서는 그러한 ‘고상관성’ 어휘를 배제해야 한다. 이를 위해 실험전에 [100만코퍼스]를 통해서 해당 어의별 공기어휘의 빈도를 측정한다.

3) 의미 중의성 해소에서 연구되는 방식은 사전적 정의를 활용한 방식과 유사하다. 이러한 점은 한 분의 심사위원에 의해서 지적된 바로 그러한 지적에 따라 2.2절에 논의가 추가되었다.

후, 매우 낮은 빈도의 공기어휘로만 실험 문항을 구성하여서 이러한 문제점을 제거하였다. 두 실험 모두 이 같은 방식을 적용하였다.

확률 추론 방식 중 본 연구에서는 베이지안 추론 확률 모형이 사용되었는데, 이 방식은 통계적인 방식을 이용한 의미 중의성 해소 알고리즘에서 많이 이용되었다(Yarowsky 1992, Gale et al. 1992). 본 연구에서는 의미구분이 된 데이터에서 해당 문맥에 대한 정보와 중의적인 단어 정보, 해당 어의의 구분자를 추출해서 활용하였다.⁴⁾ 이 연산을 위해서 본 실험에서는 [100만코퍼스]에서 해당 어의가 포함된 문맥을 문장 단위로 추출하고, 이를 토대로 해당 문맥에 포함된 공기어휘의 확률 정보와 중의적인 단어의 확률 정보, 해당 어의에 구분자가 되는 공기어휘들의 목록 및 확률 정보를 추출하였다. 이를 토대로 베이지 정리에 따라 확률을 계산하였다. 이는 3.3절에서 자세히 설명될 것이다.

3. 의미 중의성 해소 계산 모형

3.1 논리 모형

본 연구에서 활용하는 부울 논리는 부울 대수를 논리 연산에 활용하는 방식이다. 부울 논리 연산에는 여러 논리 연산자가 이용될 수 있지만, 본 연구에서는 부울 논리 연산자 AND로 검색어를 결합한 부울 검색식을 사용하게 된다. 부울 논리 연산자 AND가 생성하는 진위표는 <표 1>과 같다.

부울 논리는 핵심어의 어의에 해당 주변 어

4) 실제 실험에 활용된 알고리즘은 Manning and Schütze (2000)를 참고하였다(pp. 235-238).

<표 1> AND 부울 논리 연산 진위표

P	Q	$P \wedge Q$
T	T	T
T	F	F
F	T	F
F	F	F

휘가 공기하는 지를 AND 논리 연산으로 검색하게 된다. 만약 '물위에서 타는 배'가 {타다, 물}과 공기하게 된다면 '타다 AND 물'의 문맥에 출현하는 핵심어의 의미를 '물위에서 타는 배'로 분류하게 된다. 부울 논리식으로 이를 바꾸면 ($WORD_{타다} \cap WORD_{물}$)과 같다.

부울 논리는 Mohammad and Pedersen (2004)에서 의미 중의성 해소 알고리즘으로 활용되었는데, 코퍼스에서 중의성을 지닌 어휘의 통사적 특성들을 추출한 뒤 부울 논리를 활용해서 중의성 해소를 시도하였다. 그러나, Mohammad and Pedersen (2004)의 연구는 화자의 직관이나 주변 공기어휘를 고려하지 않고 있다는 점에서 본 연구와 차이가 있다.

김동성·최재용(2006)에서는 부울 논리 모형에 기반하여 실제 코퍼스에서 추출된 공기어휘를 의미 중의성 해소 작업에 이용하였다. 이 연구는 상대적으로 적은 양의 코퍼스를 학습 데이터로 이용해서 상대적으로 많은 양의 코퍼스에서 추출된 실험 데이터를 처리하는 부트스트래핑(bootstrapping) 방식을 특징으로 한다. 김동성·최재용(2006)은 학습 데이터에서 추출된 공기어휘의 50% 이상이 100만어절 기준으로 빈도 5 이하의 저빈도어로 구성된다는 점을 밝혔다. 그 연구에서 논리 모형에서 실제로 활용되는 어휘들은 저빈도어가 대부분이므로, 논리모형을 고려할 때에 저빈도어가 적극 고려되어야 한다는 사실이 관찰되었다.⁵⁾

이는, 앞서서도 언급한 바처럼, 심리 실험에서 고빈도어를 활용할 경우에 생기는 자료의 편향성 문제를 극복하는 방안이 된다. 또한 저빈도어가 심리실험에서 해당 어의 선택의 유용한 구분자 역할을 한다면, 논리 모형의 심리적 실재에 대한 논의에 유의미한 논거가 될 수 있다고 본다. 자세한 데이터의 빈도 반영은 4.1절에서 논의될 것이다.

3.2 확률 모형

다음 (2)의 가상 문맥에서처럼, 의미 중의성을 지닌 단어 X_A 가 A 라는 의미를 지닐 때, 해당 공기어휘 α, β 가 나타나는 환경을 고려해보자.

$$(2) \alpha \dots X_A \dots \beta$$

위의 가상 문맥의 확률적 추정은 X_A 에 대해서 결합 분포 확률(joint probabilistic distribution)인 $P(\alpha, X_A, \beta)$ 로 연산된다. 또한 그것은 아래 (3)처럼 조건부 확률로 인수분해 된다.

$$(3) P(\alpha, X_A, \beta) = P(\alpha | X_A, \beta) \cdot P(X_A | \alpha, \beta) \cdot P(\beta | X_A, \alpha)$$

이러한 확률 모형을 이용한 문맥의 추정 확률이 최대우도추정치로 연산되면 아래와 같다. 아래 (4)에서 $f(\alpha)$ 는 α 의 빈도, N 은 전체 빈도를 말한다.

$$(4) \quad a. \quad P(\alpha | X_A, \beta) = \frac{f(\alpha, X_A, \beta)}{f(X_A, \beta)} \approx \frac{f(\alpha, X_A)}{N_{Context \ X_A}}$$

5) 본 연구에서는 김동성·최재용 (2006)의 연구 결과에 근거해서 저빈도어를 활용하였다.

$$\begin{aligned}
 \text{b. } P(X_A|\alpha, \beta) &= \frac{f(X_A)}{f(\alpha, \beta)} \approx \frac{f(X_A)}{N} \\
 \text{c. } P(\beta|X_A, \alpha) &= \frac{f(\alpha, X_A, \beta)}{f(X_A, \alpha)} \approx \frac{f(X_A, \beta)}{N_{Context X_A}}
 \end{aligned}$$

(4)를 위 (3)에 적용하면 아래와 같이 간략하게 줄여질 것이다(Lapata and Lascarides 2003).

$$\begin{aligned}
 (5) \\
 P(\alpha, X_A, \beta) &\approx \frac{f(\alpha, X_A)}{N_{Context X_A}} \times \frac{f(X_A)}{N} \times \frac{f(X_A, \beta)}{N_{Context X_A}} \\
 &= P(\alpha|X_A) \times P(X_A) \times P(\beta|X_A)
 \end{aligned}$$

결과적으로 최대우도추정치는 각각의 연어들들의 확률들의 곱으로 연산된다. $n_1, n_2, n_3, \dots, n_i, \dots, n_n$ 개의 연어를 포함한 X_A 의 최대우도추정치는 $\prod_{i=1}^n P(n_i|X_A)$ 이다.

최대우도추정치는 Lapata and Lascarides (2003)에서 심리실험에 이용된 바가 있다. 이 실험에서는 British National Corpus (BNC)를 통해서 얻어진 추정치와 심리실험의 결과를 비교한 후 그 둘이 매우 유사하다는 결론을 내린다. Lapata and Lascarides (2003)에서 입증한 것은 어휘 선택 문제에서 화자 직관의 분포를 결정하는 것으로 확률 모형이 적합하다는 것이다. 이 연구에서 입증한 것은 전체 집단의 직관의 분포와 모형이 예측하는 분포가 상관성이 있다는 것이다. 본 연구에서는 이러한 점을 고려해서 집단에서의 화자 직관의 분포와 연관성이 있는 모형 중 하나로서 확률 모형을 검토하였고, 확률 모형의 여러 방식 중에서 최대우도추정치를 선택하였다. Lapata and Lascarides (2003)에서의 연구는 의미 결정과 차이가 있다. 즉, 의미적 선택 및 결정과 부합되는 모형과 집단의 직관 분포 모형은 별개로 다루어져야

한다.

최대우도추정치에는 통계적으로 자료의 빈약성(data sparseness) 문제가 있다. 즉, 최대우도추정치로 추정된 수치는 빈약한 자료들로 말미암아 신뢰성이 많이 떨어진다는 것이다. 따라서 실제 자연언어 처리에서는 최대우도추정치의 이러한 문제를 보정하는 스무딩(smoothing) 작업을 통해서 확률 추정으로 사용하고 있다(Ido et al. 1997).⁶⁾ 최대우도추정치와 관련하여 기존 자연언어 처리에서 고려된 것은 의미 결정부분에 한정되어 있다. 그러한 의미 결정은 직관의 분포와 다르며, 각각의 의미 중의성 해소 과제에 따라 다르게 결정된다. 반면 의미 결정과 아울러 본 연구에서 다룬 화자 직관의 분포는 이러한 과제별 사안에 연관되지 않고 보편적인 화자들의 의미 선택에서의 집단적 분포 특성을 나타내는 것이다.

3.3 확률 추론 모형

본 연구의 확률 추론 모형에서 고려한 베이지안 추론은 현 단계의 확률 추정치를 바탕으로 이후 단계의 확률을 추론하는 방식이다. 기본적인 베이지안 확률 결정은 아래 (6)에서 처럼 의미 중의성 해소에 활용될 수 있다. (6)에서는 단어 W 의 여러 의미 S 중 s 라는 의미를 갖는 경우를 나타낸 것이다.

$$(6) S = \underset{s \in S}{\operatorname{argmax}} \frac{P(W|s)P(s)}{P(W)}$$

베이지안 확률 결정과 관련한 본 연구에서의 실험은 해당 문맥적 정보를 가진 경우와

6) 본 연구에서는 스무딩 작업을 고려하지는 않았다.

문맥적 정보를 가지지 않은 두 가지 경우로 나뉜다. 첫 번째 실험인 [실제문맥실험]은 실제 코퍼스의 문맥을 활용한 경우로 문맥적 정보가 있는 경우이고, 두 번째 실험인 [제한문맥실험]은 가상의 제한적인 문맥에서 단어와 중의적 어휘와의 관계성을 실험한 것으로 문맥적 정보가 없는 경우이다. 자세한 사항은 4절에서 실험과 연관되어서 설명될 것이다.

문맥적 정보가 있는 경우는 나이브 베이즈 구분자(Naive Bayes classifier)를⁷⁾ 활용한 방식으로 전체 문맥의 확률을 구하였고, 문맥적 정보가 없는 경우는 베이즈 공식을 활용해서 수치를 구하였다. 나이브 베이즈 구분자를 활용한 경우에는 문맥적 정보, 해당 어의 정보, 구분자(classifier) 확률이 연산에 활용되게 된다. 이와 관련된 공식은 아래 (7)과 같다. 여기서 s 는 단어 w 의 여러 의미 S 중 하나를 말하고, a_i 는 s 의 공기어휘들 $\{a_1, a_2, \dots, a_i\}$ 를 나타내게 된다.

$$(7) P(s|a_i) = \frac{P(s) \prod_{i=1}^n P(a_i|s)}{P(a_i)}$$

단어 w 의 여러 의미 S 중 하나인 s 는 공기어휘들인 구분자 $\{a_1, a_2, \dots, a_i\}$ 가 해당 문맥 s 에 나올 확률들의 연산을 통해서 구해지는데, 해당 어의 정보인 $P(s)$ 를 고려하게 된다. 가정에 의해서 위의 (7)에서 $P(a_i)$ 는 상수이므로 실제 적용되는 수식은 (8)과 같다.

$$(8) P(s|a_i) = P(s) \prod_{i=1}^n P(a_i|s)$$

이러한 방식에 따라 어의 A 와 B 를 지나는 어휘 X 가 문맥 $\langle n_1, n_2, n_3, \dots, n_i, \dots, n_n \rangle$ 안에 있을 때, 어의 A 와 B 에 대해서 각각의 확률 $\langle n_{1A}, n_{2A}, n_{3A}, \dots, n_{iA}, \dots, n_{nA} \rangle$ 와 $\langle n_{1B}, n_{2B}, n_{3B}, \dots, n_{iB}, \dots, n_{nB} \rangle$ 을 구하게 된다. 따라서 확률 추정치는 두 개가 되며, 이를 상호 비교할 수 있다.

심리언어학적으로 연구인 Lapata and Brew (2004)는 화자직관의 분포가 아닌 화자 직관의 의미 결정 모형으로 베이지안 확률 모형이 매우 적합하다는 것을 실제 직관 실험을 통해서 입증하고 있다. 본 연구에서는 화자 직관의 분포 모형과 화자 직관의 의미 결정 모형 둘 다를 실험결과와 비교할 것이다.

4. 실험

4.1 실험 대상 및 설계

[100만코퍼스]에서 중의성을 지니는 동음이의어 중 ‘들다,’ ‘배,’ ‘신부,’ ‘다리,’ ‘거리,’ ‘막,’ ‘맞다’를 선택하였다. 이러한 어휘들의 공통점은 모두 중의적이면서도 각 어형별 어의에 따른 분포가 한 가지 어의로 크게 치우치지 않는다는 점이다. 즉, 제1어의가 전체 사용빈도 중 80%를 넘지 않는다. 예를 들어, ‘신부’의 경우 ‘카톨릭 신부’와 ‘결혼식 신부’ 두 가지의 어의가 발견되는데, 이 중 제1어 의인 ‘결혼식 신부’가 전체 사용빈도 중 80%를 넘지 않는 것을 [100만코퍼스]에서 확인하였다.⁸⁾ 마찬가지로 어의별 사용빈도가 7%이하

7) 심사자 중 한 분이 나이브 베이즈 구분자를 통한 의미 중의성 해소의 측면을 제시하였다. 이에 따라 본 절에서의 관련 논의를 추가하였다. 심사자의 지적에 감사한다.

<표 2> 실험 대상 어휘 어의 및 사용빈도

어휘	어의	사용빈도
들다	안이나 속으로 가다	55%
	위로 올리다	45%
배	신체기관	34%
	수상 운송수단	23%
	배수	43%
다리	신체기관	24%
	건축물	76%
신부	카톨릭 신부	20%
	결혼식 신부	80%
막	금방	69%
	마구	31%
맞다	옳다	55%
	물체를 몸에 받다	45%
거리	길거리	45%
	먼거리	55%

인 경우도 배제하였다.

실험의 대상으로 선택된 어휘들의 각각의 어의와 사용빈도는 <표 2>와 같다.

[100만코퍼스]에서 <표 2>에 제시된 동음이의어 중 해당 어의를 포함한 문장을 추출하였다. 3,500여개의 문장이 추출되었고, 해당 문장의 포함 어절은 평균 26.5이다. 약 92,750여개의 공기어휘 어절 타입이 수집되었다.

실험 데이터를 수집하는 과정은 다음과 같다. 첫 번째 실험인 [실제문맥실험]에서는 데

이터를 코퍼스에서 그대로 추출하되 좌우 문맥을 제한하여 문장 내 어절이 평균 5~6어절을 넘지 않도록 하였다. 문맥이 자세할수록 화자들이 너무 쉽고 빠르게 중의성을 해소할 수 있게 되므로 중의성 해소 직관이 작동하는 법을 포착하는 것이 어려워진다. 그러한 측면을 고려하여 문맥을 5~6어절로 제한한 것이다. 실험에 사용된 문장 50개를 선정하는 데는 일단은 3500여개의 문장 중 문맥이 5~6어절로 된 것을 선택한 후 그 중에서 아래 기준에 따라 47개까지 선정할 수 있었다.⁹⁾ 부족한 나머지 3개는 문맥 선정 기준을 약간 확대하여 선정하였다.

선정 기준을 좀 더 자세히 살펴보자. 논리 모형은 해당 어의가 포함되어 있는 문장에서 공기하는 어휘들을 검색한다. 따라서 이 모형은 다른 어의들 사이에 공기하는 어휘들이 서로 중복되지 않아야 한다. ‘신부’의 경우, 아래 문맥과 같이 공기어휘가 존재한다고 가정해 보자.

- (9) ㄱ. A... B... 신부(카톨릭 신부) ... C...D
- ㄴ. X... D... 신부(결혼식 신부) ... Z...B

위 (9)에서 ‘카톨릭 신부’와 ‘결혼식 신부’는 각각 {A, B, C, D}와 {X, D, Z, B}를 공기어휘로 하는데, 중복되지 않는 어휘만을 고려하면 {A, C}와 {X, Z}¹⁰⁾이 될 것이다. 이러한

8) 강범모(2005)의 연구에 따르면 사전에서 발견되는 다수의 동음이의어의 경우 제1어의 사용빈도가 전체의 약 98%를 상회한다고 한다. 이러한 동음이의의 경우 실제 중의성 해소 작업이 필요가 없는 바, 이런 어휘를 제거하기 위해서 [100만코퍼스]에서 사용빈도를 조사하였다.

9) 문장을 선정하는 방식은 벡터 공간 모델(vector space model)을 사용하였다. 이러한 작업은 어휘 자체의 빈도와 여러 문맥에서의 빈도도 고려하기 위함이다.

10) 실제로 고려된 어휘는 명사, 형용사, 동사, 부사와 같은 내용어(content word)로 형식어(functional word)를 고려하지 않았다.

서로 중복되지 않는 어휘목록을 이용해서 ‘카톨릭 신부’의 경우에는 $(WORD_A \cap WORD_C)$ 를 ‘결혼식 신부’의 경우에는 $(WORD_X \cap WORD_Z)$ 를 검색식으로 이용한다. <표 2>에 나와 있는 어의가 포함된 문장을 모두 추출해서 해당 어의에 공기하는 모든 어휘들을 추출한 다음, 서로 중복되지 않는 어휘집합을 만들어 냈다. 실험에 사용한 문장으로는 기본적으로 어의 별 공기어휘가 한 개 이상 포함된 것을 선정하였다. 한 개 이상 포함 조건이 만족이 되지 않은 경우에는, 해당 어의에만 공기하는 어휘가 포함이 되지 않지만 중복이 되는 어휘들로 이루어지도록 하였다. 이것은 확률 모형과 확률 추론 모형을 만족하기 위한 것이다. 실제 어휘들이 너무 높은 확률로 나타나는 경우에는 자료의 치우침 현상이 문제가 되기 때문이다.¹¹⁾ 예를 들어, (7)에서 {A, B}나 {X, Z}가 해당 어의에 너무 높은 빈도로 출현해서 실제 확률에서 높게 계산된다면, 자료의 편향성이 문제가 된다. 실험 문장 설정에서는 이러한 편향성 문제를 극복하기 위해서 공기어휘들의 출현 빈도가 [100만코퍼스]에서 20이하인 경우만을 고려하였다. 출현 빈도에 따른 확률로 계산을 하면 최대 0.009, 최저 0.0000001로 매우 낮은 경우이다.

두 번째 실험인 [문맥제한실험]에서는 공기어휘와 해당 어의와의 관계를 측정하고자 하였다. 이 실험의 경우 논리 모형을 완전히 배제하기 위해서, 추출한 공기어휘 중 해당 단어의 어의들에 중복되는 경우만을 고려하였다. 따라서 위의 (9)에서와 같은 경우에는 {B, D}만을 고려하였다. 이 경우에 {B, D}의 출현빈도와 예상빈도를 고려하여 예상빈도와 출현빈

도 사이에 너무 많은 차이가 없는 경우를 선택하였다. 실험에 사용된 어휘들의 실제 빈도는 이 경우에 매우 다양하게 나타나는데, 빈도는 [100만코퍼스] 기준으로 324에서 1 사이이다. 이 경우에 출현 빈도를 기준으로 한 확률은 최대 0.0005, 최저 0.0000001이다. 최대 빈도어인 324 출현 빈도어인 경우에는 예상빈도가 3.59이다. 그리고 최저 빈도어의 경우에는 출현 빈도는 1이고 예상빈도는 0.01이다. 개략적으로 60% 이상의 어휘들이 예상빈도와 출현빈도의 비율이 약 1000:1 정도로 낮은 경우가 많은 것을 감안하여, 예상빈도가 너무 낮은 경우를 제외한 것이다. 또, 예상빈도가 너무 높은 경우, 예를 들면 출현 빈도는 1인데, 예상빈도는 0.1로 약 출현 빈도와 예상 빈도가 약 1:10 정도의 경우도 제외하였다. 이것은 첫 번째 실험인 [실제문맥실험]에서 자료의 편향성 문제를 해결하고자 노력과 일치한다.

4.2 실험 절차

본 연구에서는 실험참가자가 웹페이지에 접속해서 실험에 응할 수 있도록 온라인 형태의 실험을 실시하였다. 각 실험참가자별 선택 결과는 웹 서버에 연결된 DB에 기록되도록 하였고, 그 결과물은 DB형태로 저장되도록 하였다.¹²⁾ 실험은 크게 예비 실험과 본 실험으로 구성되었다. 예비실험은 본 실험에 앞서서 기술적인 문제와 사용자 인터페이스 및 실험의 편리성과 같은 문제점을 파악하기 위하여 실시하였다. 예비실험은 서울소재 대학 대학원생 7명이 참여하였고, 본 실험에서는 학부생 2~4학년 33명이 참여하였다. 두 실험 모두 자

11) 자료 편향성과 관련된 자세한 논의는 Lapata and Lascarides (2003)을 참조.

12) 페이지 주소는 <http://corpus.mireene.com/test.php>.

발적인 참여로 이루어진 무급 실험이었다.¹³⁾

예비 실험은 온라인상에서 개인별로 실시하도록 한 후, 실험 후 참여자들이 참여후기를 작성하도록 하였다. 참여후기를 통해 실험상의 문제점을 파악하고 이를 본 실험에 반영하였다. 이어 본 실험은 동일 장소에서 동일 시간대에 실시하여 실험 환경을 제한하였다. 실험 전에 설명문건을 나누어주고 실험절차에 대해 간략히 설명한 후에 실험에 들어갔다.

실험의 내용을 살펴보면, 실험은 두 개의 부분으로 구성되었고, 각 부분별로 50개씩 문항이 제시되었다. 4.1에서 설명한 바와 같이, 첫 번째 실험은 [실제문맥실험]으로 5~6개 정도의 어절로 구성이 된 문장들을 코퍼스에서 추출된 형태대로 제시하였다. 두 번째 실험은 [제한문맥실험]으로 코퍼스에서 추출된 공기어휘를 다음 (10)의 분열문 구조 X에, 그리고 핵심어휘를 Y에 배치한 문장 형태로 제시하였다. 이는 제한적인 문맥환경에서 실제로 화자의 직관에 부합하는 것이 어떤 모델인지를 밝히기 위한 것으로, 부울 논리 모형이 적용될 수 없는 환경을 만들기 위한 목적에서 구상되었다. (부록 [제한문맥 실험 문항]의 21번)

(10) ㄱ. X하는 것은 Y이다.

ㄴ. 떨어지는 것은 거리이다.

위 (10)은 실제 실험에 사용된 문항으로 실험참가자에게 ‘떨어지다’라는 단어와 중심어인 ‘거리’를 분열문 구조에서 제시하였다. 실험참가자는 (8)의 문장을 읽고 ‘거리’가 ‘먼거리’와 ‘길거리’ 중 하나를 선택하였다. 논리 모형은 하나의 어의에만 공기하는 어휘로 문맥을 검

13) 실험 후 실험에 참여한 사람들에게는 소정의 기념품이 제공되었다.

색하는 방식이다. 따라서 ‘거리’의 경우 ‘먼거리’의 어의를 선택하는 경우는 ‘먼거리’와만 공기하는 어휘나 ‘길거리’를 선택하는 경우에는 ‘길거리’에만 공기하는 어휘가 문맥에 존재하는 경우에 가능하다. 위 (10)에서 제시된 ‘떨어지다’라는 단어는 ‘길거리’와 ‘먼거리’의 어의에 모두 공기하는 어휘로 부울 논리 모형의 적용이 불가능하다. 이러한 경우에 실험참가자들의 어의 선택 모형을 측정하는 것이 이 실험의 목적이었다.

예비 실험에서 문항별로 10초의 시간을 주었으나, 실험 시간이 너무 짧다는 지적이 많았으므로, 본 실험에서는 시간을 20초로 늘렸다. 각 문항별로 화면상에 20초 초읽기(countdown)가 있어서, 20초의 시간이 지나면 자동으로 다음 문항으로 넘어가도록 하였다. 또는 실험참가자가 20초의 시간 전에 어의를 선택하는 경우에도 바로 다음 문항으로 넘어가도록 하였다.¹⁴⁾

실제 실험은 부록에 제시된 문항을 웹페이지에 하나씩 보여주고, 문항에서 질문하는 단어의 어의를 나열하였다. 실험참가자가 어의를 하나 선택하면, 다음 문항으로 넘어 가도록 하였다. 아래 (11)은 문항별 구성의 예이다.

(11) 신부가 의미하는 뜻을 고르시오.¹⁵⁾

[신부]도 예쁘고 저 친구도 예쁘구나
 카톨릭 신부 결혼식 신부

14) 반응 속도가 중요한 변수일 수 있으나 본 연구는 반응 속도 자체보다는 어의 선택 모형 선택에 주안점을 두었으므로 예비실험을 통해 드러난 실제 실험자들의 요구를 수용하여 응답시간 20초를 부여하였다.

15) 본 문항은 [실제문맥 실험 문항] 49번이다.

McNamar (2005)에 따르면 화자 직관을 이용한 의미 분류 실험에서 심리언어학적으로 의미적 점화(semantic priming) 효과가 매우 중요하게 작용한다고 한다. 이 효과는 의미적으로 연관이 있는 어휘들간에 연상작용이 일어난다는 것으로 school-class라는 단어쌍은 school-house라는 단어쌍 보다 활성화(activation)가 더 빠르게 나타나는 것을 설명하게 된다. 만약 본 실험에서 의미적 점화 효과가 영향을 미친다면 실험에 대한 요인으로 작용할 가능성이 많다. 하지만 본 실험은 의미적 점화 효과 요인을 관찰하기 위한 실험이 아니므로, 최대한도로 실험의 요인이 되지 않도록 억제할 필요가 있었다. 따라서 의미 분류 문항에 영향을 미치지 않도록 분류하는 문항을 무작위로 배치하였다. 예를 들면 처음 문장이 ‘배’의 의미를 분류하는 문장이면 다음 문장은 ‘다리’를 분류하는 문장을 배치하였다. 그리고, 이러한 점은 예비 실험을 통해서 검증하였다.

마지막으로 소수점을 나타내는 마침표를 포함한 문장 내 문장부호(?, ‘!’“”) 및 기타 부호(#@*&...)를 모두 삭제하였다. 이것은 문장 부호나 기타 부호가 문맥에 미치는 영향을 없애기 위한 조치였다. 실험 설문 문항은 부록에 기재될 것이다.

5. 실험 결과

5.1 실제문맥 실험

본 소절에서는 응답자의 응답 신뢰도 측정을 위한 카파통계 적용, 문항별 응답자간 일치도 결과, 실험 결과와 모형별 측정치와의 상관관계, 그리고 계산 모형별 정확도에 대해

여 차례대로 논하기로 한다.

실험결과에 대한 신뢰성 검증을 위해 카파통계를 이용하였다. 즉, 실험참가자별 어의 선택이 우연의 일치인지 아니면 실제 신뢰할만한 선택인지에 대한 확률적 타당성을 검증하였다(Cohen 1960). 이 통계는 여러 실험참가자가 관여된 언어학 실험에서 실험자별 의미 분류 결과가 신뢰할 만한 것인지를 검증하는 것이다(Carletta 1995, Passonneau 2004). 주관적인 분류실험에는 우연한 일치와 신뢰할 만한 일치가 혼재되어 있다는 가정을 바탕으로, 카파통계는 이 중 우연한 일치를 제외한 결과로서 신뢰할 만한 일치를 연산해 내는 방식이다.¹⁶⁾

$$(12) \kappa = \frac{P_a - P_e}{1 - P_e}$$

위에서 P_a 는 실험참가자별 응답내용간의 우연한 일치를 뜻하고 P_e 는 일치비율을 의미한다. 이 통계는 수치적으로 0.6이상이면 신뢰 가능하고, 0.8이상이면 매우 신뢰할 만한 수준을 나타낸다(Cohen 1960, Krippendorff 1980). 이 실험의 경우에 P_a 는 0.94이고 P_e 는 0.50으로 κ 는 0.88로 측정되었다. 이는 매우 신뢰할만한 수준으로 실험참가자의 응답일치는 우연한 일치가 아닌 의미 있는 결과로 평가된다.

이번에는 응답자간 일치도를 살펴보기로 하자. 첫 번째 실험인 [실제문맥실험]의 결과는 각 문항당 98%이상의 실험참가자가 하나의 어의에 일치하는 현상을 나타내었다. 문항별 응답결과간 일치도를 보면, 50문항 중 49문항

16) Bruce and Wiebe (1998)에서는 이 방식을 활용해서 실제 화자 직관을 이용한 데이터를 통해서 일치와 불일치 데이터를 통계적으로 검증하여 재분류를 하는 방식을 논의하였다.

에서 응답자들간 일치도가 98%에 이른다.¹⁷⁾

실험참가자들간 문항별 일치 비율은 [100만 코퍼스]에서 측정된 결과와 비교되었다. 예를 들어 아래 (10)에서 응답자들은 모두 ‘거리’라는 중의적 표현에 대하여 ‘길거리’라는 어의를 선택하였다.

(13) 거리는 여전히 활기에 넘쳐 있었다.¹⁸⁾

이러한 실험결과와 각 계산 모형별 측정치와 차례로 비교해 보기로 한다. 논리 모형은 ‘거리’의 경우 {여전히, 활기, 넘치다}의 공기어휘가 각각의 어의에서 중복되지 않은 경우를 조사하게 된다. 위 (13)의 경우 2개의 어휘 ‘활기’와 ‘넘치다’가 ‘길거리’의 경우에만 공기하는 어휘들이므로 ‘길거리’는 확률적으로 0.67이고, 또 다른 어의 ‘먼거리’는 중복되는 경우를 포함하지 않으므로 0.00이 된다.

확률 모형은 공기어휘가 해당 어의로 사용될 확률을 최대우도추정치를 통해서 구하게 된다. 위의 (13)에서 공기하는 어휘인 {여전히, 활기, 넘치다}가 ‘길거리’라는 어의에 공기하는 확률들과 ‘먼거리’라는 어의에 공기하는 확률들을 구한다. 이 확률을 (5)에서 가정대로 독립가정을 근거로 최대우도추정치를 곱하여 산출하게 된다.

확률 추론 모형의 경우에는 [실제문맥실험]의 경우에는 문맥에서 등장하는 베이즈 확률을 이용한 것이므로, 문맥 상 확률을 반영하기 위해서 (8)에서 제시된 공식에 의거해서 나이브 베이즈 확률을 구하였다. 이 확률 추정치는 (13)의 예에서는 {여전히, 활기, 넘치다}의

각각의 확률을 나이브 베이즈 공식에 의해서 (8)에서 제시된 나이브 베이즈 공식으로 값을 구한 다음 이를 서로 곱한다.

이러한 계산을 바탕으로 모형별로 산출된 측정치와 실험결과사이의 피어슨 상관계수를 측정하였다. 피어슨 상관계수는 0.4이상이면 비교적 높은 상관관계를 나타내며 수치가 크면 클수록 더 높은 상관성을 나타낸다(김홍규 1990: 235). 여기서 사용된 비교 추정치는 상관계수의 결정계수 (coefficient of determination)인 r^2 값을 제공한 r^2 을 사용하는데, 관계된 두 변인의 공유변량 (shared variance) 혹은 공통변량 (common variance)의 비율을 나타내게 된다. 논리 모형의 경우 $r^2=0.415$ ($N=300, p<0.01$),¹⁹⁾ 최대우도추정치를 통한 확률 모형의 경우 $r^2=0.338$ ($N=300, p<0.01$), 베이지안 확률을 통한 확률 추론 모형의 경우 $r^2=0.096$ ($N=300, p<0.01$)로 측정이 되었다.²⁰⁾ 따라서 논리 모형이 가장 상관성이 높은 것으로 측정되었다. 이러한 상관계수의 통계적 유의성을 t -검증을 통해서 살펴보았는데, 각각이 $N=300$, 유의확률 $p<0.05$ 일 경우에 논리 모형의 경우 $t=10.870$, 확률 모형의 경우 $t=9.55$, 확률 추론 모형의 경우 $t=3.349$ 로 산출되었다. 따라서, 논리 모델의 신뢰도가 상대적으로 가장 높은 것으로 나타났다.

이어서, 실험결과에서 우세한 결과를 해당 어의로 간주하고 이를 해당 모형별 결정 방식

17) 실험결과에 대한 통계는 다음과 같다. 표준편차 0.2919, 분산 0.008, 첨도 7.342, 왜도 2.983 이다.

18) 실제문맥 실험 문항 1번.

19) N은 모집단 개수로서, 어의가 3개인 경우 각 어의별로 0 또는 1을 취한다고 볼 때, 전체개수가 150개(50문항 * 3개 어의)가 되므로, 이를 두 배로 하면 (실험결과, 비교 모형) 합계 300개가 된다.

20) 피어슨 상관계수는 SPSS v. 12.0 패키지를 이용하여 측정하였다.

과도 비교해 보았다. 예를 들어, 실험결과 어의 A와 어의 B가 98:2로 나타났더라도 아직은 중의성이 해소된 것이 아니다. 98:2를 근거로 어의 A를 선택하는 절차, 즉 결정 절차가 필요하다. 실험에서 98:2로 나온 경우는 일단 어의 A로 결정된 것으로 가정하였다. 그렇다면 각 계산 모형별로는 각기 산출된 측정치를 바탕으로 어떤 결정을 하게 될 것인가? 그리고 그러한 결정이 실험결과에서의 결정과 어느 정도 일치하는가? 이러한 점에서 실험에서의 결정 결과와 모형별 결정 결과간 일치도를 측정해 보았다.²¹⁾ 일치도가 높을수록 정확도가 높다고 할 수 있다. 결과적으로 논리 모형은 이 경우에 95% 정도의 일치를 보이고 있었으며, 확률 모형은 75%, 확률 추론 모형은 90%의 일치를 보였다.

지금까지의 논의를 정리해 보면, 논리 모형이 세 모형 중 실험결과에 가장 상관성이 높으며, 또한 정확도가 제일 높은 것으로 평가되었다. 확률 추론 모형은 상관성이 상대적으로 낮지만, 정확한 결정을 하는 것으로 평가되었으며, 확률 모형은 실제 실험결과와 상관성이 높지만, 어의 결정에서는 정확도가 떨어진다는 점이 밝혀졌다.

5.2 제한문맥 실험

본 소절에서는 전절에서와 마찬가지로, 응

21) 만약 실험결과가 어의 A가 우세하고, 확률 모형에서 어의 B가 확률적으로 높아서 어의 B로 결정되었다면 이 경우에는 불일치의 경우로 판단이 된다. 이러한 방식으로 확률 추론 모형은 확률이 높은 것으로 결정하고 이를 서로 비교해 보았다. 논리 모형은 해당 문맥에서 관련된 어의를 가장 많이 포함한 경우를 해당 어의로 선택한다.

답자의 응답 신뢰도, 실험 결과와 모형별 측정치와의 상관관계, 그리고 모형별 정확도를 논한다. 문항별 응답자간 일치도가 다양해서 일치수준에 특별한 의미를 부여하기 어려우므로 논의에서 제외한다.

[제한문맥실험] 결과에다가 카파 통계를 적용, 실험자의 일치 신뢰도를 측정하였다. 수치는 0.65정도로 신뢰할 만한 수준이었다.

[제한문맥실험]은 부울 논리 모형을 제외한 환경으로 제시된 것이어서, 나머지 두개의 계산 모형, 즉 확률 모형과 확률 추론 모형과 비교되었다. 확률 추론 모형은 문맥적인 요소가 없이 분열문만으로 제한된 문맥에서 공기 어휘와 중심어와의 관계성을 나타내는 구조로, (6)과 같은 베이즈 공식에 의해 결과를 산출하여 비교하였다.

두 번째 실험도 첫 번째 실험과 마찬가지로 피어슨 상관계수 측정을 통해서 상관성을 측정하였다. 확률 모형의 경우 결정계수인 $r^2=0.433$ ($N=300$, $p<0.01$)이고, 확률 추론 모형의 경우는 결정계수가 $r^2=0.073$ ($N=300$, $p<0.01$)으로 측정되었다. 확률 모형의 경우는 확률 추론 모형의 경우보다 상관성이 높은 것으로 나타났다. 확률 추론의 경우에는 낮은 상관관계로 고려될 만한 수준은 아니었다(김홍규 1990: 235). 앞 절에서와 마찬가지로 상관계수의 통계적 유의성을 t -검증을 통해서 살펴보았는데, 각각이 $N=300$, 유의확률 $p<0.05$ 일 경우에 확률 모형의 경우 $t=10.631$, 확률 추론 모형의 경우 $t=3.406$ 으로 산출되었다. 따라서, 확률 모형의 신뢰도가 상대적으로 가장 높은 것으로 나타났다.

두 번째 실험도 첫 번째 실험에서와 같이 가장 우세한 결과를 해당 결정 모델과 비교해 보았다. 이 경우에 확률 추론 모형은 약 85%

정도 일치되는 것을 보였지만 확률 모형은 약 60%정도가 일치되는 것으로 나타났다.

실험결과를 상관성의 측면에서 검토해본 결과, 확률 모형이 실험에서 직관의 분포와 비교적 높은 상관관계를 나타낸 반면에, 확률 추론 모형은 상관성이 거의 없는 경우를 보였다. 그러나, 결정 모형에서는 확률 추론 모형이 상당히 정확한 것으로 나타났다.

5.3 결과 토의

본 연구에서 고찰된 것은 의미 중의성 해소 계산 모형간에 처리의 순서화가 가능한가라는 의문이다. 먼저, 첫 번째 실험인 [실제문맥실험]은 기준에 따라 코퍼스에서 추출된 문장을 그대로 화자에게 실험한 것이다. 자료의 편향성을 배제한 추출 문장을 대상으로 한 실험에서, 화자들의 직관은 98%이상 일치하였다. 고빈도 공기 어휘로 인해 생길 수 있는 자료의 편향성 문제를 피하기 위해서 본 실험에서는 코퍼스에서 매우 낮은 확률로 출현하는 어휘들로 추출된 문맥을 실험하였다. 본 논문에서의 가설은 매우 적은 정보를 가진 문맥에서도 화자 직관은 논리 모형을 근거한 의미 선택이 가능하다는 것이었다. 실제 문맥 실험 결과는 그러한 가설에 부합한다. 이 실험에서 화자들의 직관 분포와 가장 유사한 상관성을 갖는 경우는 논리 모형이었다. 그리고 화자 직관에 근거한 결정 모형에서도 논리 모형은 정확하게 예측하는 것을 알 수 있었다. 따라서 화자의 직관모형을 계산적으로 설명하는 방식에서 부울 논리 모형은 매우 정확하게 예측하고 결정하는 것을 알 수 있었다.

5.1절에서 논의한 바처럼, [실제문맥실험]에서 화자들 간 98%의 일치도를 보이는 반면

전혀 예외적인 경우가 하나결과가 나왔다. 구체적으로, 50개의 문항 중 아래 (14)에 제시된 하나의 문항에서만 응답이 크게 엇갈렸다. 응답자중 45%는 첫 번째 어의인 ‘신체기관’을, 55%는 두 번째 어의인 ‘수상운송수단’을 선택하였다.

(14) 영환이 배를 바라본다.²²⁾

(14)의 공기어휘 집합을 보면 {영환, 바라보다}이다. 그런데 이러한 공기어휘들은 ‘신체기관’의 경우에도 또 ‘수상운송수단’이란 어의와도 공기하는 것으로 코퍼스에서 발견된다. 이처럼 주어진 공기어휘 집합이 각각의 어의에 중복되어 있는 경우에는 논리 모형이 적용될 수 없다. 그런데 이 경우 확률 모형을 살펴본 결과 43%대 57%로, 실험 결과와 매우 유사한 수치가 발견되었다. 이러한 결과를 보이는 경우가 전체 실험 문항에서 하나밖에 없었으므로, 논리 모델의 적용이 불가능하면 확률 모델이 우세하게 적용된다는 것을 단정적으로 말할 수는 없었다. 따라서 이러한 관찰이 어느 정도 보편성을 띠고 있는지 확인하기 위한 다른 형태의 실험이 필요하였다.

두 번째 실험은 논리 모형을 제외하고 확률 모형과 확률 추론 모형을 비교하는 방식으로 설계되었다. 이 방식에서 화자의 직관을 살펴보면 확률 모형이 확률 추론 모형에 비해서 높은 상관성을 보이고 있었다. 그러므로 논리 모형이 적용되지 못하는 문맥에서는 확률 모형이 높은 상관성을 지님을 알 수 있었다.

확률 추론 모형은 직관의 분포와는 상이한 것으로 관찰되었지만, 화자 직관에 근거한 결

22) 실제문맥 실험 문항 33번.

정 모형에서는 상당히 정확하게 일치함을 알 수 있었다. 이것은 확률 추론 모형이 의미 중의성 해소에서 널리 사용되는 확률 추론 방식임을 입증하는 결과로 볼 수 있다(Ide and Vernois 1998). 즉, 결정 모형은 직관의 분포를 예측하는 모형이 아닌, 어의를 결정하는 모형이므로, 추론 모델은 확률 추론 모델은 매우 정확하게 어의를 결정할 수 있었다. 이러한 결정 모형은 결정 직관과 부합하는 결과를 나타낼 수 있지만, 화자 직관의 분포와는 정확히 일치하지 않는 것을 알 수 있었다. 그리고, 확률 모형이 상대적으로 제일 부정확한 결정 모형임이 밝혀졌는데, 이것은 자료의 빈약성(data sparseness)이 문제가 되는 경우엔 최대우도추정치가 신뢰성이 낮은 방법이라는 기존의 연구 결과와 부합된다.

화자의 직관이 논리 모형과 유사성이 많음이 관찰되었는바, 의미 중의성 해소를 위한 기계적 알고리즘을 개발할 때 이를 구체적으로 구현할 필요가 있다. 또한 기존 연구와의 비교도 좀 더 치밀하고 체계적으로 이루어질 필요가 있을 것이다. 이러한 점들은 추후 연구로 미룬다.

참고문헌

- [1] 강범모 (2005). 동음이의어의 사용양상. **어학연구**, 제41권 1호, pp. 1-29.
- [2] 김동성·최재용 (2006) 저빈도어를 고려한 개념학습 기반 의미 중의성 해소. **언어와 정보**, 제10권 1호, pp. 21-46.
- [3] 김홍규 (1990). **사회통계분석**. 나남.
- [4] Ahlswede, T. (1995) Word sense disambiguation by human subjects: computational and psycholinguistic implications. Workshop on Acquisition of Lexical Knowledge from Text, 31st Annual Meeting of the Association for Computational Linguistics, June 21-26, 1993, pp. 1-9.
- [5] Bruce R. and J. Wiebe. (1998) Word sense distinguishability and inter-coder agreement. In *Proceedings of 3rd Empirical Methods in Natural Language Processing*, Granada. ACL SIGDAT, pp. 53-60.
- [6] Carletta J. (1996) Assessing agreement on classification tasks: the kappa statistic. *Computational Linguistics*, Vol 22, pp. 249-254.
- [7] Ido D., L. Lee, and F. Pereira. (1997) Similarity-Based Methods For Word Sense Disambiguation. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and Eighth Conference of the European Chapter of the Association for Computational Linguistics*.
- [8] Cohen, J. (1960) A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, Vol. 20, pp. 37-46.
- [9] Gale W., K. Church, and D. Yarowsky. (1992) Estimating upper and lower bounds on the performance of word-sense disambiguation programs. 30th Annual Meeting of the Association for Computational Linguistics, University of Delaware, Newark, DE, July. Association for Computational Linguistics, pp. 249-256.
- [10] Ide N. and J. Vernois. (1998) Introduction to the special issue on word sense disambiguation: the state of the art. *Computational Linguistics*,

- Vol. 24, pp. 1-40.
- [11] Krippendorff. K. (1980) *Content Analysis*. Beverly Hills, CA: Sage Publications.
- [12] Harris, Z. (1964) Distributional structure. In Fodor, Jerry and Jerrold Katz (ed.): *The Structure of Language*, Pritence-Hall, pp. 33-49.
- [13] Jackson P. and I. Moulinier. (2002) *Natural Language Processing for Online Application*. John Benjamins Publishing.
- [14] Lapata M. and A. Lascarides. (2003) A probabilistic account of logical metonymy. *Computational Linguistics*, Vol. 29, pp. 261-315.
- [15] Lapata, M. and C. Brew (2004) Verb class disambiguation using informative priors. *Computational Linguistics*, Vol. 30, pp. 45-73.
- [16] Manning C. and H. Schütze. (2000) *Foundations of Statistical Natural Language Processing*. MIT Press.
- [17] Mihalcea R. and D. Moldovan. (2004) An Automatic Method for Generating Sense Tagged Corpora. In *Proceedings of the American Association for Artificial Intelligence (AAAI 1999)*, Orlando, FL, July 1999, Vol. 30, pp. 45-73.
- [18] Mohammad S. and T. Pedersen. (2004) Combining lexical and syntactic features for supervised word sense disambiguation. In the *Proceedings of the Conference on Computational Natural Language Learning (CoNLL)*, Boston, MA , May, pp. 6-7.
- [19] McNamara, T. (2005) *Semantic Priming: Perspectives from Memory and Word Recognition*. Taylor & Francis Corp.
- [20] Passonneau R. (2004) Computing reliability for coreference annotation. In *Proceedings of LREC*, Lisbon.
- [21] Vernois J. (1998) A study of polysemy judgements and inter-annotator agreement. Programme and advanced papers of the Senseval workshop, Herstmonceux Castle, England, pp. 2-4.
- [22] Yarowsky D. (1992) Word sense disambiguation using statistical models of Roget's categories trained on large corpora. In *Proceedings of the 14th International Conference on Computational Linguistics, COLING'92*, Nantes, France, August, pp. 454-460.

1 차원고접수: 2006. 7. 18

2 차원고접수: 2006. 10. 9

최종게재승인: 2006. 11. 28

부 록

실제 실험에 사용한 문항은 아래와 같다. 아래 문항에서 [] 안에 들어 있는 중심어를 포함한 문맥과 해당 중심어의 어의를 제시하고 제시된 어의 중 하나를 선택하도록 하였다. 중심어별 어의는 <표 2>에 나와 있다.

[실제문맥실험] 문항

1. [거리]는 여전히 활기에 넘쳐 있었다
2. 내려서자마자 하천을 가로지르는 야트막한 [다리]로 이어졌다
3. [들어] 있다
4. [거리]에는 온통 가을이 물결치고 있었다
5. 자 한잔 [들]개나
6. 아침이었는데 태양이 [막] 떠오르고 있었습니다
7. 올 가을 [거리]엔 또 어떤 유행 컬러가 휩쓸까
8. 아이가 [막] 출구를 빠져 나가는 순간이었다
9. 은행 앞이 노랗게 물들어 간 [거리]를
10. 그도 [막] 잠에서 깨어난 모양이었습니다
11. 부리는 어두운 잿빛 남색이고 [다리]는 암적색이다
12. 그러면 [막] 날개짓을 웅얼한다
13. [다리] 달달 떨지 말라고
14. [맞아] 불문과야
15. 페리호가 부산항이 한눈에 들어오는 [거리]까지 벗어났다
16. 까부니까 [맞]았지
17. 장군님은 점잖게 [배]에 올라탔습니다
18. 아니 [배]는 왜 탕지
19. 게이턴호수로 [배]가 들어섰다
20. 해방 이래 [거리]마다 정치 폭력이 날뛰었다
21. 박홍 [신부]께서 - 우리 학교 송자도 그랬잖아
22. 또 [들어] 있네
23. 그러나 [배]는 이미 떠나가고 있었다
24. [신부] 친구는 일절을 열창했습니다
25. 귀향할 [배]를 기다리게 하여 주었다
26. 눈앞이 어질어질하고 [다리]에 힘이 빠졌다
27. [다리]가 떨리고 있었다
28. [막] 돈이 들어온데요
29. 오징어는 [다리]를 잡아당기어 내장을 빼낸다
30. 정말 [배]가 오는군
31. [다리] 난간에 붙은 쇠조각조각 뜯어 갔다
32. 워커발로 [막] 걷어차기도 하고요
33. 영환이 [배]를 바라본다
34. 텔런트 홍학표 [신부]
35. [다리]를 구부려 보고 돌려도 보자
36. 그라모 딱 [맞다]심더
37. 아이고, 우리 새택 [신부]께서 일을 저지르셨구먼
38. 예 [맞다]지요 아파
39. 대학로 [거리]에는 여러 번 나갔지요
40. 9[배] 액수는
41. 1. 6[배] 61.87 퍼센트
42. 도둑 [들라]
43. [배] 안 아프고
44. 작은아버지 [드]세요
45. 그 몇 [배]라고
46. 그럼 사과나 [드]실래요
47. 1의 5[배]
48. [신부]도 예쁘고 저 친구도 예쁘구나
49. 어서 [배]에 오르십시오
50. 아이고 [배] 아프다

[제한문맥실험] 문항

1. 만나게 되는 것은 [거리]이다
2. [다리]가 사용되는 것은 배이다
3. [거리]를 만드는 것은 사람이다
4. [들]게 되는 것은 마음이다
5. [막] 만들어지는 것이 말이다
6. 보여지는 것은 [거리]이다
7. [막] 만들어지는 것이 친구이다
8. 낮은 것은 [거리]이다
9. [다리]가 사용되는 것은 물이다
10. [막] 다가오는 것이 사람이다
11. [다리]가 사용되는 것은 개이다
12. 통하는 것이 [맞다]
13. 결국 [다리]가 사용된다
14. [맞]게 되는 것이 세상이다
15. [거리]를 만드는 것은 시간이다
16. [배]가 이용되는 것이 최근이다
17. 사회가 드러나는 것이 [배]이다
18. 알려진 것은 [다리]이다
19. [맞]게 되는 것이 시간이다
20. 통하는 것은 [배]이다
21. 멀어지는 것은 [거리]이다
22. 장군이 사용하는 것은 [배]이다
23. 물건이 보이는 것이 [배]이다
24. 물이 사용되는 것은 [배]이다
25. 오히려 [다리]가 사용된다
26. [들]게 하는 것이 한국인이다
27. [막] 다가오는 것이 공부이다
28. 같게 되는 것은 [맞다]
29. [신부]가 들어오는 것은 집이다
30. 일정한 것은 [거리]이다
31. 지나가는 것이 [들다]
32. [다리]가 밀린다
33. [막] 오는 것이 친구이다.
34. [막] 다가오는 것이 느낌이다
35. 기대되는 것이 [들다]
36. 그렇다는 것이 [맞다]
37. 많은 것은 [거리]이다
38. [맞]게 되는 것이 미국이다
39. [막] 다가오는 것이 움직인다.
40. [신부]가 생각나는 것은 발이다
41. 주는 것이 [맞다]
42. 보는 것은 [신부]이다
43. [들]게 되는 것은 생각이다
44. 부르는 것은 [신부]이다
45. [들]게 하는 것이 집안일이다
46. 신경이 보이는 것이 [배]이다
47. [신부]가 생각나는 것은 목사이다
48. 구성되는 것이 [들다]
49. 통하는 것은 [신부]이다
50. [맞]는 것이 현실이다