

□ 기술개설 □

한국어 의미 해석시 중의성 해소에 대한 연구†

한국과학기술원 조정미* · 김길창**

● 목	차 ●
1. 서 론	3.2 의미격 분류
2. 의미 해석의 방법론	3.3 대표 조사 분류
2.1 규칙 기반 방법	3.4 한국어 동사 격률 표현 방법의 표준화
2.2 예문 기반 방법	4. 신경망을 이용한 의미 중의성 해소 방법
2.3 통계적 방법	4.1 자료 표현
2.4 신경망을 이용한 방법	4.2 신경망을 이용한 단어 중의성 해소
3. 한국어 의미 중의성 해소를 위한 지식 기반 구축	4.3 신경망을 이용한 의미격 결정
3.1 명사 및 동사 분류 체계	5. 결 론

1. 서 론

한국어를 비롯한 모든 자연 언어는 중의성을 가지고 있다. 따라서 자연 언어 처리는 이러한 중의성을 해소해 나가는 과정이라고 말할 수 있다. 중의성은 언어 처리의 단계에 따라 여러 단계로 분류할 수 있다. 본 논문에서는 한국어 의미 해석시 발생하는 의미 중의성을 대상으로 이를 해소하는 방안을 제시하고자 한다.

의미 해석시 발생하는 중의성은 크게 두 가지로 볼 수 있다.

첫째, 의미 중의성을 지닌 단어에 의해 발생하는 중의성이다. 일상 생활에서 흔히 쓰이는 단어 중에는 문맥이나 상황 정보에 의해서만 그 단어의 의미가 결정되는 것이 많이 있다. 예를 들면, “그는 눈을 들어 하늘을 보았다”와 같은 문장에는 ‘눈’과 ‘들다’라는 의미 중의성을 지닌 단어가 사용되었으며, 이러한 단어들

에 대한 정확한 의미 결정 없이는 이 문장을 정확하게 해석할 수 없다. 의미 중의성을 지닌 단어의 의미는 문장에서 함께 나타난 다른 단어들에 의해서만 결정된다.

둘째, 단어와 단어 사이의 관계에서 발생하는 중의성이다. 단어와 단어 사이의 관계는 형태, 의미, 기능이라는 문법적 기준 가운데 어떠한 기준을 중심으로 하느냐에 따라 구문 관계와 의미 관계로 나눌 수 있다. 구문 관계는 문장에서의 격조사에 의해 표현되며 문장에서의 문법적 기능을 의미한다. 의미 관계는 일반적으로 ‘의미격’이라 하여 서술어에 대한 구성 성분의 역할을 나타낸다. 예를 들면, “그는 망치로 못을 박다”와 “그는 콩으로 매주를 쑤다”라는 두 문장에서 성분 ‘망치’와 ‘콩’은 모두 조사 ‘로/으로’에 의해 부사적 기능을 나타낸다. 그러나 그들이 서술어에 대해 갖는 역할은 도구의 역할과 재료의 역할로 구분된다. 이러한 의미격은 조사 정보뿐만 아니라 서술어와 구성 성분의 의미를 모두 고려하여야 결정할 수 있다.

의미 중의성 해결에 대한 연구는 주로 영어를 대상으로 하여 규칙 기반 방법, 예문 기반

† 본 연구는 과학재단의 목적 기초 과제 “한국어 이해에 나타나는 중의성 문제 처리 모델에 관한 연구”의 부분 지원을 받은 것입니다.

*학생회원

**중신회원

방법, 통계적 방법, 신경망을 이용한 방법 등 다양하게 연구되어 왔다. 본 논문에서는 먼저 이와 같은 다양한 방법론들에 대해 살펴보고 한국어 처리시 발생하는 의미 중의성에 대한 해소 방안으로 신경망을 이용한 중의성 해소 방법을 제시하고자 한다.

신경망을 이용한 시스템은 학습 과정을 통하여 입력 자료의 특징을 스스로 추출하는 능력과 이들을 일반화하여 규칙을 획득하는 능력을 가지고 있어 결과의 객관성을 보장할 수 있다. 또한 문제 적응 능력이 뛰어나 학습하지 않은 자료에 대해서도 만족할 만한 결과를 보이며, 또한 학습을 통하여 규칙을 쉽게 확장시킬 수 있다. 이러한 특징은 규칙의 모호성이 강하고, 입력 형태에 변형이 많으며 처리해야 할 정보가 많은 자연 언어 처리 분야에 적합하다.

2. 의미 해석의 방법론

의미 해석시 발생하는 의미 중의성은 문장 안에서 관계를 갖는 단어와 단어사이의 제한 정보에 의해 대부분 해소된다. 그리고 이러한 단어간 관계 정보를 어떻게 구성하고 적용하느냐에 따라 다양한 방법론들이 제시되었다.

규칙 기반 방법이란 단어간의 관계 제한 정보를 언어학적 지식에 기반하여 잘 정의된 규칙으로 만들어 적용하고자 하는 것이며, 예문 기반 방법이란 관계 제한 정보를 직접 예들로 표현한 것이다. 통계적 방법이란 단어간의 제한 관계 정보를 대량의 말뭉치를 이용하여 자동적으로 추출하여 적용하며, 신경망을 이용한 방법은 대량의 학습 자료와 신경망을 통해 단어간의 관계를 학습한다. 각 방법론들에 대해 좀더 자세히 살펴보자.

2.1 규칙 기반 방법

규칙을 기반으로 하는 방법은 의미 해석 방법 중 가장 오래되고 일반적인 방법이다. 이 방법에서는 선택 제한(selectional restriction) 기법을 포함하여 여러 규칙들과 격틀(case frame) 정보, 동사와 명사의 의미 정보를 이용하여 문장의 구성 성분들간의 관계를 분석한다.

선택 제한이란 문장에서 특정 관계를 가질 수 있는 단어들간의 제한 정보를 의미한다. 예를 들면, 어떤 동작을 취하는 동사들마다 필요한 성분이 있고 이러한 성분이 될 수 있는 단어의 의미 조건에 제한이 있다는 것이다. 즉, '먹다'라는 동사는 먹는 주체가 되는 행위자 성분과 먹는 대상이 되는 대상 성분을 가지며, 행위자 성분으로 나타날 수 있는 단어의 의미 조건은 '생물체'여야 하며, 대상 성분으로는 '먹을 수 있는 무엇'이어야 한다. 이러한 선택 제한을 이용하여 문장의 구조를 분석해 내거나 전치사구 부착 문제 등을 해결한다.

선택 제한 기법은 단어의 의미 조건에 정확히 일치할 것을 요구한다. 따라서 격틀이나 단어 제한 정보로 명확히 명시된 것에 대해서는 정확한 결과를 내 줄 수 있으나 그렇지 않은 것에 대해서는 결과를 내주지 못한다. 또한 시스템의 확장을 위해서는 규칙이나 의미 정보의 개선이나 확장이 매우 어렵다. 따라서 이러한 방법은 제한된 영역에서는 좋은 결과를 얻을 수 있으나 영역을 확장시키고자 할 경우 어려움이 있다.

규칙 기반한 의미 중의성 해소에 대한 기존의 연구를 살펴보자. [22, 5, 17, 8, 19, 7]은 선택 제한 기법에 의한 단어간의 다양한 제한 정보와 선호 규칙(preference rule)을 이용하여 의미 해석시 발생하는 의미 중의성을 해소하고자 하는 연구들이다. [13]은 구문 정보와 약간의 의미 정보를 이용하여 문장의 격 구조를 해석하는 모델을 제안하였다.

2.2 예문 기반 방법

[20, 11]은 예문에 기반한 언어 처리의 적용 실험과 전망을 제시했다. 지식으로 규칙을 사용하는 규칙 기반 방법은 시스템의 확장이 어렵고 특정한 영역의 정보의 사용이 어려우며, 정확한 일치를 요구하므로 실패하는 경우가 많다. 이러한 문제를 극복하고자 예문들의 데이터베이스를 이용하였다. 이 방법은 시스템의 확장이 쉽고 특정 영역의 정보를 사용할 수 있게 하고, best 일치를 하므로 실패하지 않는 견고한 결과를 출력할 수 있다. 입력의 문장이 들어오면 이 문장과 예문이 들어 있는 데이터

베이스의 내용과의 일치 여부를 검사를 통해 어떠한 예에 속하는지 가려지게 된다. 일치 여부는 시소러스 등을 이용한 의미 유사도에 의한 거리 계산 방법을 이용한다.

예문에 기반한 방법은 best 일치를 이용하므로 어떠한 문장이라도 분석 결과를 구할 수 있어서 매우 견고하다. 또한 예문이 많을수록 분석할 수 있는 범위가 늘어나므로 시스템의 확장도 용이하다. 그러나 문장과 예문간의 의미 유사도 계산과 구조 일치 정도를 측정할 수 있는 함수에 매우 의존적이다. 그리고 예문들의 효율적인 저장과 검색 방법 등에 대한 연구가 병행되어야 한다.

2.3 통계적 방법

통계적 방법은 대량의 말뭉치로부터 단어간의 관계 정보를 자동으로 추출하고, 이들을 다양한 평가함수로 평가하여 중의성을 해소한다.

[4]는 서로 관계를 맺는 단어간의 관계의 정도를 정보 이론의 개념을 이용한 식으로 평가하는 것을 제시하였다. 이 논문에서는 단어 사이에 발생하는 관계에 대하여 word association ratio를 구한다. 이 수치는 객관적인 관점에서 개념 단어와 개념 단어간의 관계와 개념 단어와 기능 단어간의 관계를 고려하여 많은 양의 말뭉치를 통해 얻은 것으로 association ratio는 아래와 같이 정의하였다. 이 측정치는 단어간 연관 정도를 나타내며 언어 처리의 여러 분야에서 유용하게 사용되고 있다.

$$I(x, y) = \log_2 \frac{P(x, y)}{P(x)P(y)}$$

x와 y는 관계를 갖는 각 단어를 의미한다. 그리고 P(x,y)는 단어 x와 y가 한 문장 또는 정해진 간격 내에서 같이 나타나는 확률이고 P(x)와 P(y)는 각 단어가 나타나는 확률이다. 말뭉치의 크기가 클수록 신뢰성 있는 확률이 나온다. [1, 14, 12]는 단어간 관계 제한 정보를 대량의 말뭉치로부터 자동으로 학습하고자 하였다. [2]는 의미 중의성 해소를 위한 규칙을 변환 기반 오류 보정 학습법(Transformation-Based Error-Driven)에 의해 말뭉치로부터 자동 습득하여 적용하는 모델을 제안하였으

며, [18, 23, 24]에서는 말뭉치와 MRD(machine readable dictionary)를 이용한 단어 중의성 해소에 대한 모델을 제안하였으며, [15, 16, 3, 6]은 말뭉치로부터의 문맥 정보를 이용한 의미 중의성 해소에 대한 모델을 제안하였다.

2.4 신경망을 이용한 방법

규칙을 이용한 방법과 예문을 이용한 방법은 분석 과정에서 인간이 정의하는 많은 정보를 필요로 한다. 그러나 신경망을 이용한 방법은 학습에 의해 많은 정보를 추론해내므로 처리과정의 객관성을 보장할 수 있고, 또한 필요한 정보의 양을 줄일 수도 있다. 몇 가지 신경망을 이용한 의미 해석 시스템에 대해 살펴보자.

[10]에서는 형태소에서 discourse 단계까지의 정보를 집적한 memory network을 이용하여 중의성 문제를 해결하는 모델을 제안하였다. [21]은 전치사구 부착의 문제를 명사구에서 해소하기 위하여 의미적인 관계의 표현을 사용한 back propagation network과 의미적인 제한 조건과 구문적인 제한 조건을 집적한 표현을 이용한 relaxation network으로 구성된 네트워크 구조를 이용하였다. [9, 25]는 영한 기계번역에서 전치사구의 올바른 한국어 번역을 위해 WordNet과 같은 On-line Lexical Database와 back propagation network를 이용한다.

3. 한국어 의미 중의성 해소를 위한 지식 기반 구축

의미 해석시 발생하는 의미 중의성 해소에는 여러 가지 언어적 지식 기반이 필요하다. 이러한 지식 기반은 전문가에 의해 구축될 수도 있으며, 말뭉치로부터 자동으로 구축될 수도 있다. 전문가에 의한 구축된 지식은 정확한 반면 그 구축이나 확장에 어려움이 있으며, 자동으로 구축된 지식은 지식의 구축이나 확장이 용이한 반면 지식에 오류가 포함될 가능성이 높다.

다음의 내용들은 단어의 의미 중의성 해소나 단어간 관계 해석을 위해 필요한 일반적인 지

식 기반에 대한 것이다.

3.1 명사 및 동사 분류 체계

의미 해석에서 사용되는 대표적인 지식으로는 의미 분류 체계가 있다. 영어에 대해서는 이러한 의미 분류에 대한 연구가 많이 진행되어 WordNet과 같은 On-line Lexical Database도 구축되어 있다. 그러나 한국어에 대한 의미 분류 연구는 아직 부족한 형편이다.

의미 분류는 비교적 언어간 차이가 없다. 그러나 분류가 세분화될수록 그 언어의 특징이 나타나고, 분류의 하위 개념인 개별 단어에서는 언어간 차이가 확실히 드러난다. 또한 의미 중의성을 지닌 단어와 같은 경우는 그 언어에 의존적일 수밖에 없으므로 특정 언어에 맞게 만들어진 의미 분류 체계를 그대로 번역해서 다른 언어 영역에 적용한다고 하는 것은 어느 정도 한계가 있다.

일반적으로 의미 정보를 저장하는 자료 구조로는 나무 구조(tree structure)를 이용한다. 이 구조에서 각 노드는 하나의 의미를 표현한다. 나무 구조 안에는 상위 개념과 하위 개념이 존재하는데, 상위 개념은 보다 포괄적인 개념을 나타내고 하위 개념은 상위 개념의 세분화된 개념을 나타낸다. 따라서 하위 개념은 상위 개념의 속성을 모두 따르게 된다.

본 논문에서의 명사 분류 체계는 기존의 여러 연구에서 연구되어진 분류 체계를 종합하여 의미 중의성 해소에 적합하게 재구성하였다.

동사는 중심 의미를 고려한 어휘 의미론적인 방법에 따라 그 체계를 수립하였다[26]. 어휘 의미론적 방법이란, 여러 어휘 중 공통되는 동사류의 등가의 의미를 설정하고 자료의 조건에 따라 상호 관계를 결정하는 방법이다. 이와 같은 분류법은 동사 의미에 따른 격들 설계에 매우 유리하다. 어휘 의미론적 분류 방법의 가정은 다음과 같다

첫째, 동사는 동작동사와 상태동사의 두 범주로 나뉜다. 동작 동사란 종래의 동사를, 상태 동사란 종래의 형용사를 의미한다.

둘째, 동작동사는 '무엇이 어찌한다, 누가 어찌한다'의 검증법으로 판정하고 상태동사란 '무엇이 어떠하다, 누가 어떠하다'라는 검증법

으로 판정한다.

셋째, 동작동사, 상태동사의 하위 범주는 동치 의미를 고려하여 해당 동사류의 중심 의미에 따라 결정한다. 동작동사는 수여 의미, 소유 의미, 이동 의미, 부정 의미, 비교 의미, 상접 의미, 조종 의미, 보류 의미, 가변 의미로 다시 세분화되고 이들은 다시 하위 개념으로 세분화된다. 상태동사는 도량 의미와 평결 의미로 나뉘고, 도량 의미는 기본 도량, 부차 도량으로, 평결 의미는 추론 의미와 감각 의미로 다시 세분화된다.

본 논문은 기존의 한국어에 대한 의미 분류에 대한 연구들을 기본으로 하여 의미 중의성 해결을 위한 관점을 반영하여 의미 분류를 하였다. 명사와 동사의 분류 체계는 부록 A, B을 참고하기 바란다.

3.2 의미격 분류

의미격 분류는 단어와 단어사이에 나타날 수 있는 의미 관계를 정의하는 것이다. 본 논문에서는 의미격을 서술어에 대한 구성 성분의 역할로 정의하였다. 격 정의를 위해서는 구성 성분의 문장 내 발생 빈도와 중요성을 고려하였다. 자주 나오거나 문장 구성에 필수적인 성분은 세밀하게 분류하고, 출현 빈도가 적거나 필수적이지 않은 요소는 어느 정도 통합하여 개략적으로 분류한다. 이 때 문장의 중심 의미는 정확히 나타내고 그 외의 추가적인 의미는 개략적 의미 관계로 나타낸다. 설정된 격은 모두 30개이다. 분류를 위해 여러 선행 연구를 참조하였으며 일상적인 한국어 문장을 대상으로 하여 잘 나타나지 않는 격은 제외하였으며, 자주 발생하거나 문장의 의미를 정확히 표현하기 위해 구별해야 하는 것, 예를 들면 도구격, 재료격, 수단격 등은 모두 구별하도록 하였다. 이와 같은 방법으로 새로이 제안된 격 분류는 표 1과 같다.

3.3 대표 조사 분류

한국어 조사에는 격조사, 보조사 및 복합 조사가 있다. 한국어 문장은 영어와 달리 조사에 의하여 구문적 역할이 결정되는데, 이것은 의미격을 제한하는 효과가 있다. 즉, 조사 '로부

표 1 의미격 분류

의미격	정의	예
행위자격(agent)	의지가 있는 행위의 주체	소녀가 꽃을 꺾었다.
경험자격(experiencer)	의지가 없는 행위의 주체	그녀는 슬프다.
대상격(object)	행위의 대상	아이가 유리창을 깬다.
사동주격(causer)	사동주	어머니가 옷을 입히다.
묘사격(characterized)	형용사에 의해 성격이 규정되는 대상	하늘이 파랗다.
정의격(isa)	명사에 의해 성격이 규정되는 대상	침대는 가구이다.
여격(dative)	수신자 및 수여동사의 수용자	그가 나에게 책을 준다.
제공자격(supplier)	송신자 및 수여동사의 제공자	나는 그에게 돈을 꾸었다.
수익적(benefactive)	이익이나 손해를 받는 대상	그는 나라에 공헌했다.
행위내용격(act cut)	동사의 내용을 표현	소년은 열심히 공부를 했다.
도구격(instrument)	행위의 도구 역할	칼로 두부를 썬다.
재료격(material)	행위의 재료 역할	벽돌로 집을 지었다.
역할격(role)	역할, 자격, 지위	그는 선배로서 충고했다.
수단격(mean)	행위의 수단이나 방법	그들은 영어로 말했다.
이유격(reason)	사건이나 행동의 원인, 이유	친구는 감기로 결석했다.
용도격(purpose)	사건이나 행동의 목적, 용도	꽃을 선물로 주었다.
짝격(partner)	행위를 함께 해야하는 대상	그녀는 비행사와 결혼했다.
동반자격(accompany)	공동 행위자	그는 강아지와 공원에 갔다.
비교격(comparative)	비교의 상대가 되는 대상	그는 형보다 크다.
관계격(relative)	대상과 관계를 맺는 대상	피란색이 너에게 어울린다.
변성격(transformed)	변화의 도달점이 되는 상태나 대상	그는 의사가 되었다.
장소격(loc at)	행위나 사건이 일어난 장소	그는 운동장에서 논다.
시작장소격(loc from)	출발점이 되는 장소	그는 학교에서 떠났다.
도달장소격(loc to)	도달 장소, 목적지	선생님이 미국에 도착했다.
시간격(time at)	행위나 사건이 일어난 시간	내일 만나자.
시작시간격(time from)	출발점이 되는 시간	이침부터 미가 온다.
도달시간격(time to)	도달 시간	저녁까지 일을 해야 한다.
양태격(manner)	동작이 이루어지는 양상이니 방법	게미처럼 열심히 일하라.
정도격(quanty)	수량, 정도, 빈도	그는 10km를 걸었다.
장면격(scope)	동작의 장면 및 기술 범위	꿈에서 만나자.

터'로 나타나는 성분은 서술어에 대해 시작장소, 시작시간, 재료의 역할 중 하나만을 선택할 수 있다. 따라서 본 논문에서는 의미격을 효율적으로 결정하기 위해서는 조사를 의미격에 대응시켜 이용하는 방법이 고려되었다.

그러나 실질적으로 조사를 이용하여 구문적 성질을 분석하는 작업에는 많은 어려움이 따른다. 특히 보조사의 경우는 다양한 격으로 자유롭게 사용되기 때문에 이들의 구문 성질을 규정하는 것은 매우 어렵다. 보조사는 특정한 격

으로 특징짓지 않고 격 결정과 관계없는 특징 의미만 더해주므로, 다른 조사와의 분류를 위해 하나로 묶어 분류할 뿐 세분화하지 않았다. 그러나 예외적으로 격 결정에 뚜렷한 기여를 하는 몇 개의 보조사가 있는데, 이들은 격조사처럼 각각 따로 분류하였다. 복합조사란 두 개 이상의 조사로 이루어진 조사로서, 격조사-격조사 결합, 격조사-보조사 결합, 보조사-보조사 결합 등 3가지 유형이 있다. 이들 유형 중 격조사가 들어 있는 복합 조사는 그 격조사에 따라, 격조사가 두 개 이상일 경우에는 마지막에 위치한 격조사에 따라 분류하였다.

예를 들면, 복합조사 '에서가'는 '가'로 간주한다. 보조사-보조사 결합은 보조사로 분류하였다. 예를 들면 '마저도'는 보조사로 간주한다. 한국어 문장에서는 특히 조사 생략이 빈번히 발생한다. 조사가 생략되었을 경우에는 기본적으로 조사 정보를 가지지 않지만, 생략 현상을 나타내기 위하여 따로 분류하였다.

조사는 매우 많은 종류가 있으므로 본 논문에서는 이를 사용 용법과 격 결정 유형에 따라 23가지의 대표 조사를 정의하고 이들에 따라 조사를 분류하여 사용하였다. 조사 분류는 표 2와 같다.

표 2 대표 조사 분류

대표 조사	그 외 해당 조사	가질 수 있는 격의 종류
가	이, 에서가	행위자격, 경험자격, 묘사격, 대상격, 사동주격, 도구격, 이유격, 반성격
께서	께서, 께서는	행위자격, 경험자격, 묘사격, 대상격, 사동주격
를	을	대상격, 행위자격, 시좌장소격, 장소격, 도달장소격
에		장소격, 시간격, 역할격, 수익격, 장면격, 관계격, 도달장소격, 정도격, 여격, 제공자격
에서		장소격, 시작장소격, 장면격
에다	에다가	장소격
에게	께, 한테, 한테다가, 에게다가	여격, 제공자격, 행위자격, 수익격
로	으로	도달장소격, 역할격, 재료격, 도구격, 수단격, 용도격, 이유격
로써	으로써	도구격, 재료격, 도달시간격
로서	으로서	역할격, 정도격
보다		비교격
만큼		정도격, 비교격
와	과	짝격, 동반자격, 비교격
따나		양태격
처럼	같이	양태격, 비교격
에게서	한테서	제공자격
로부터	으로부터	시작장소격, 시작시간격, 재료격
서부터	에서부터	시작장소격, 시작시간격, 재료격
부터		시작장소격, 시작시간격
까지		도달장소격, 도달시간격
더러	한테, 한테다가, 보고	여격
보조사	은, 도, 마지, 조차, 마다	모든 격
생략		모든 격

3.4 한국어 동사 격률 표현 방법의 표준화

한국어는 언어 분류학상 첨가어에 속한다. 첨가어는 성분에 첨가되는 요소에 의하여 그 성분의 역할이 결정되므로 첨가 요소의 역할이 중요하다. 이와 함께 한국어의 동사는 어떠한 첨가 요소를 가지는 성분을 필요로 한다. 동사 격률이란 문장에서 특정 관계를 가질 수 있는 서술어와 구성 성분간의 제한 정보를 명시한 것을 의미한다. 즉, 어떤 동작을 취하는 동사들마다 필요로 하는 성분이 있고 이러한 성분이 될 수 있는 단어의 의미 조건에 제한을 나타낸다. 따라서 격률을 설계하기 위해서는 의미격에 대한 정의와 단어의 의미 조건에 대한 표현 방법이 정해져야 한다. 단어의 의미 조건은 분류 체계와 같은 언어적 지식을 이용할 수도 있고, 직접 예문으로 의미 조건을 대신할 수도 있다. 다음의 예에서 1)은 분류 체계를 이용한 격률의 표현이고 2)는 예문을 이용하여 의미 조건을 나타낸 격률 표현의 예이다.

- 1) [떡다[행위자격 : 생명체] [대상격 : 식료]]
- 2) [떡다[행위자격 : [아이, 강아지, 꽃, 영화]] [대상격 : [우유, 뼈다귀, 물, 사과]]]

격률 설계를 위한 의미격과 의미 조건 표현 방법이 결정되었다고 하더라도 동사 격률을 설계하기란 어려운 일이다. 왜냐하면 설계해야 하는 동사의 수가 많을 뿐 아니라 의미 중의성을 지닌 동사 같은 경우는 각각의 의미마다 적합한 격률을 정의해야 하기 때문이다. 그리고 격률에 표현할 의미 조건의 범위를 결정하는 것 역시 매우 어려운 일이다. 따라서 근래에는 이런 격률을 반자동으로 추출할 수 있는 방법에 대한 연구와 신경망을 이용한 학습으로 격률을 대신하는 연구가 진행 중이다.

4. 신경망을 이용한 의미 중의성 해소 방법

그림 1은 본 논문에서 제안한 신경망을 이용한 모델의 의미 중의성 해소 과정을 나타낸 것이다. 먼저 구문 분석 결과로부터 동사와 명사, 조사 쌍을 입력으로 받아 이를 신경망에서 처

리 가능한 형태의 자료 표현으로 변환한다. 그런데 명사들 중에는 의미 중의성을 지닌 명사들이 많다. 그리고 입력 표현은 이와 같은 복합적인 의미를 모두 표현하게 된다. 이 복합적인 의미들 중 입력문에서 해당하지 않는 의미 속성들은 격을 결정하는데 커다란 장애 요인이 된다. 따라서 이를 해결하기 위해 단어의 중의성 해소 단계를 거치게 된다. 마지막으로 조사와 중의성이 해소된 명사 의미 속성 및 동사 의미 속성을 이용하여 주어진 문장의 상황에 적합한 의미격을 결정하게 된다.

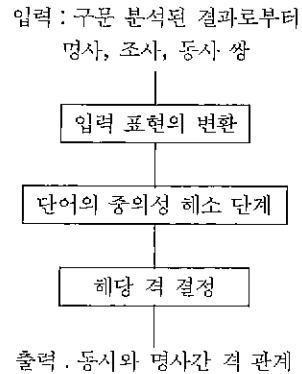


그림 1 의미 중의성 해소 과정

4.1 자료 표현

일반적으로 신경망은 기호로 표현된 자료들을 다루지 못한다. 그러나 자연 언어 처리에 사용되는 자료들은 대부분 기호이므로 이들을 신경망에서 처리하려면 먼저 숫자 형태로 바꾸어야 한다. 신경망에서의 자료 표현 방법으로는 분산 표현 방법과 국부 표현 방법이 있다. 분산 표현 방법은 단어 등과 같은 하나의 개념을 신경망 내의 여러 노드들로 분해하여 표현하고, 국부 표현 방법은 하나의 개념을 하나의 노드에 대응시킨다. 국부 표현은 분산 표현에 비해 간단하고 이해하기 쉬운 장점이 있지만 단어의 의미를 표현하는 경우와 같이 표현해야 할 개념의 수가 많을 경우에는 신경망의 노드를 너무 많이 필요로 하기 때문에 실질적으로 표현이 불가능하다. 따라서 표현할 개념의 특징에 따라 국부 표현과 분산 표현 중 알맞은 방법을 선택하여야 한다. 일반적으로 자연 언

어 처리 분야에서는 분산 표현 방법이 많이 사용되고 있다.

본 논문에서는 국부 표현과 분산 표현을 모두 사용한다. 신경망의 최종 출력 노드는 의미 분석한 후 결정되는 격에 대응되는데 이는 그 수가 많지 않기 때문에 국부 표현 방법을 사용한다. 또한 신경망 입력의 하나인 조사도 국부 표현 방법으로 표현한다. 본 신경망에서는 표 1과 표 2에서 언급한 의미격과 조사 분류를 이용한다. 주요 입력 정보인 명사와 동사의 의미 정보 표현은 앞에서 언급한 이유로 하여 각각 분산 표현 방법을 사용한다. 이를 효과적으로 표현하기 위해서는 일반적으로 잘 정의된 미세 자질(micro feature)이 정의되어야 하는데 본 논문에서는 이를 위해서 명사의 의미 속성과 동사의 의미 속성들을 정의하여 사용한다.

명사와 동사 의미의 분산 표현을 위해서 의미 정보를 분산 표현으로 변환하는 과정이 필요하다. 먼저 명사나 동사 사전으로부터 각 단어에 대한 의미 분류 코드를 구한다. 이러한 단어의 의미 분류 코드 각각을 의미 단위로 하여 분산적으로 표현하는 것도 가능하다. 그러나 의미 분류 코드는 모든 단어의 의미 분류를 위해 정의한 것이기 때문에 중복되거나 격 분석에 불필요한 정보들을 많이 포함하고 있으므로 이들을 보다 유용하고 정선된 형태로 바꾸어 표현하는 것이 바람직하다[25]. 분산 의미 표현의 단위가 되는 이러한 의미 형태를 의미 속성이라고 부른다. 단어의 의미 속성을 구하려면, 먼저 단어의 의미를 표현할 수 있는 의미 속성 집합을 정의하고, 단어 분류 코드와 의미 속성간의 변환을 위한 변환 함수를 정의하여야 한다. 본 논문에서 정의하여 사용한 명사와 동사의 의미 속성 집합과 변환 함수는 부록 C를 참고하기 바란다

4.2 신경망을 이용한 단어 중의성 해소

우리들은 일상 생활에서 의미 중의성을 지닌 단어들을 많이 사용하고 있다. 예를 들면, '눈'이라는 단어는 'snow'와 'eye'라는 두 가지 의미를 지니고 있다. 그러나 우리는 이러한 단어가 사용되는 상황에 의해 어느 한 가지 의미만으로 받아들인다.

따라서 단어로부터 의미 속성을 추출하는 과정에서는 단어가 가진 모든 의미의 의미 속성을 모두 구한다. 왜냐하면 단어 자체만을 보고서는 어떠한 의미로 쓰일지 알 수 없기 때문이다. 그러나 일단 문장에서 다른 단어들과 함께 사용될 경우는 여러 의미 중 어느 한 의미로 결정하여야 한다.

단어 중의성 해소 신경망은 단어가 가질 수 있는 의미들 중에서 실제로 문장 안에서 사용된 의미만을 뽑아 낸다. 이러한 작업을 해야 하는 이유는 단어의 의미 중의성이 격의 올바른 결정을 방해하기 때문이다.

단어 중의성 해소 신경망은 명사, 동사의 의미 분류 코드와 조사 정보를 문맥 정보로 이용하여 문제가 되는 단어의 중의성을 해결하였다. 즉, 명사의 의미 중의성은 조사와 동사를, 동사의 의미 중의성은 명사와 조사를 이용하여 해결하였다. 예를 들면 명사 '눈'은 기상 현상 물로서의 의미 속성과 신체 일부로서의 의미 속성을 모두 지니는 의미 중의성을 지닌 단어이다. 문장 "눈이 내린다"에서는 주격 조사 '이'와 동사 '내린다'의 의미 속성을 이용하여 '눈'의 의미를 기상 현상물로 결정할 수 있다. 또 다른 문장 "눈이 아프다"에서는 주격 조사 '이'와 형용사 '아프다'의 의미 속성을 이용하여 '눈'의 의미를 신체의 일부로 결정할 수 있다.

중의성 해소 신경망은 명사 중의성 해소 신경망과 동사 중의성 해소 신경망으로 나뉘며, 모두 하나의 은닉층을 갖는 back propagation network이다. 이들은 입력으로 명사의 모든 의미 분류에 대한 분산 의미 표현과 조사의 국부 표현, 동사의 분산 의미 표현을 받아 출력으로 각각 중의성이 해소된 명사 분산 의미 표현과 동사 의미 표현을 내준다. 명사의 분산 의미 표현과 동사의 분산 의미 표현은 각각 30개와 40개의 노드를 가지고 조사의 국부 표현은 23개의 노드이다. 따라서 단어 중의성 해소 신경망의 입력 노드는 총 93개이고 출력 노드는 명사 중의성 해소 신경망의 경우는 30개, 동사 중의성 해소 신경망의 경우는 40개이다. 학습의 효율을 높이기 위해 명사 중의성 해소 신경망의 경우 동사는 문맥에 따라 의미 중의

성이 해소된 분산 의미 표현을 학습 자료의 입력으로 이용하며, 동사 중의성 해소 신경망의 경우 명사는 중의성이 해소된 분산 의미 표현을 자료 입력으로 이용하였다. 학습을 통해 입력으로부터 각각 명사와 동사의 단어 중의성을 해소하여 문맥과 어울리는 의미 속성만을 추출하며 단일 의미인 단어는 입력 표현을 그대로 출력한다.

4.3 신경망을 이용한 의미적 결정

격 결정 신경망은 단어 중의성 해소 신경망의 출력을 바탕으로 하여 각 단어들의 문장에서의 격을 결정한다. 이 신경망 역시 back propagation network으로 입력층과 출력층, 하나의 은닉층을 갖는다. 입력층의 노드는 23개의 이진 조사 코드와 중의성 해소 신경망의 결과로 얻은 30개의 명사 의미 속성, 40개의 동사 의미 속성이며 출력층은 30개의 격에 대한 가능값(plausible value)이다. 출력층의 30개 격에 대한 가능값을 비교하여 가장 큰 것을 선택함으로써 의미적을 결정하게 된다. 신경망의 출력으로 나타나는 의미적은 국부 표현방식으로 표현된다.

이 신경망은 “동사와 명사의 의미 정보와 조사만으로 격을 결정할 수 있다”라는 기본 가정을 하고 있다. 이는 신경망을 이용하여 어떠한 문제를 해결하고자 할 경우 입력의 구조를 미리 결정하여야 하는데, 격 분석의 대상인 문장 전체를 입력으로 하려면 대상 문장을 매우 정규화된 단위로 제한하여야 하는 문제를 해결하기 위한 것이다. 이렇게 문장을 명사와 동사의 단위로 나누어 각각에 대한 분석을 함으로써

본 격 결정 신경망은 어떠한 형태의 문장이라도 분석이 가능하다. 그러나 실질적으로는 격 결정을 위해 단어의 위치 정보나 문장 내의 다른 문맥 정보와 같은 문장 정보를 필요로 하는 경우가 발생하고, 이러한 경우 본 신경망 모델의 확장이 필요하다.

5. 결 론

본 논문에서는 의미 해석시 발생하는 의미 중의성 해소에 대한 다양한 방법론들을 살펴보고 한국어 의미 중의성 해소 방법으로 신경망을 이용한 모델을 제안하였다. 신경망을 이용한 시스템은 규칙의 모호성이 강하고 입력 형태에 변형이 많으며 처리해야 할 정보가 많은 의미 중의성 해소에 유용하게 이용될 수 있다. 또한 학습하지 않은 자료에 대해서도 어느 정도의 결과를 보장할 수 있다.

본 논문에서 제안한 신경망을 이용한 의미 해석 모델은 문장의 전체를 고려하지 않는다. 즉 “동사와 명사의 의미 정보와 조사만으로 격을 결정할 수 있다”라는 기본 가정을 하고 있다. 그러나 격 결정을 위해 명사, 조사, 동사뿐만 아니라 다른 문장 정보가 필요한 경우가 있다. 따라서 문장 전체를 고려할 수 있는 확장된 신경망 모델에 대한 연구가 더 필요하다.

그리고 본 논문은 문장에 서술어가 한 번 나오는 단문의 경우를 대상으로 하고 있다. 그러나 단문 이외의 내포문이나 복합문의 경우, 문장의 서술어와 구성 성분간의 관계 뿐만 아니라 명사 수식어와 명사와의 관계, 동사와 동사(중심 동사와 의존 동사)간의 관계를 모두 고려하여야 한다. 이를 위해서는 명사 수식어와 명사, 동사와 동사 사이의 격을 해석할 수 있는 모델이 필요하다. 이것은 본 논문에서 제안한 신경망 구조를 그대로 이용하고 입력 형태를 약간 수정함으로써 쉽게 구현할 수 있다.

또한 신경망 모델과 다른 방법을 결합하여 서로의 단점을 보완할 수 있는 모델에 대한 연구도 필요하다. 신경망 모델은 초기 가중치 결정이 전체 학습 결과에 미치는 영향이 크다. 따라서 통계적 방법을 이용하여 신경망 모델의 초기 가중치를 어느 정도 객관적 자료에 근거

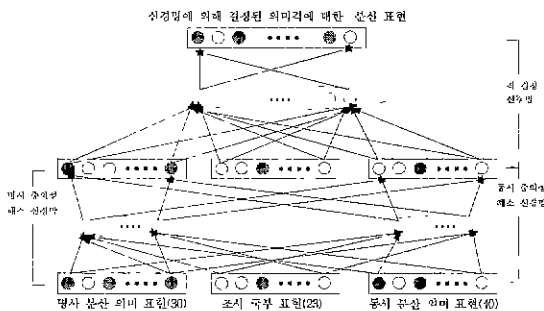


그림 2 신경망 구조도

하여 결정하는 것은 의미 중의성 해소의 결과나 학습 소요 시간 등에 긍정적인 영향을 미칠 것이다. 또는 전체 의미 해석 시스템에서 학습이 필요한 단계에서만 부분적으로 신경망 모델을 적용하는 모델도 어느 한 방법론에 의한 것보다 더 정확하고 빠른 결과를 보일 것이다.

참고문헌

- [1] M. R. Brent, "Automatic Acquisition of Subcategorization Frames from Untagged Text," ACL, pp. 209-214, 1991.
- [2] E. Brill, A Corpus-Based Approach to Language Learning, Ph.D. Thesis, Univ. of Pennsylvania, 1993.
- [3] J. M. Cho and G. C. Kim, "Korean Verb Senses Disambiguation Using Distributional Information from Corpora," NLPRS, Vol.2, pp. 691-696, 1995.
- [4] K. Church and P. Hanks, "Word Association Norms, Mutual Information, and Lexicography," ACL, 1989.
- [5] Don Fass, "Collative Semantics," COLING, 1986.
- [6] F. Fukumoto and J. Tsuchi, "Automatic Recognition of Verbal Polysemy," COLING, 1994.
- [7] J. R. Hobbs and J. Bear, "Two Principles of Parse Preference," COLING, 1990.
- [8] X. Huang, "Semantic Analysis in XTRA," Computers and Translation, 1988(3).
- [9] W. S. Kang, Jungyun Seo, K. S. Choi, and G. C. Kim, "A Neural Network Method for the Semantic Analysis of Prepositional Phrases in English-Korean Machine Translation," Computer Processing of Chinese and Oriental Language, 1994.
- [10] H. Kitano, H. Tomabechi, and L. Levin, "Ambiguity Resolution in DMTRANS PLUS," ACL, 1989.
- [11] S. Kurohashi and M. Nagao, "A Method of Case Structure Analysis for Japanese Sentences Based on Examples in Case Frame Dictionary," IEICE TRANS. INF. & SYST., Vol. E77-D, No.2, pp.277-239, 1994.
- [12] H. Li and N. Abe, "Generalizing Case Frames Using a Thesaurus and the MDL Principle," Recent Advances in Natural Language Processing, pp.239-248, 1995.
- [13] Rey-Long Liu and Von-Wun Soo, "An Empirical Study on Thematic Knowledge Acquisition based on Syntactic Clues and Heuristics," ACL, 1993.
- [14] C. D. Manning, "Automatic Acquisition of A Large Subcategorization Dictionary from Corpora," ACL, pp. 235-242, 1993.
- [15] F. Pereira and N. Tishby, "Distributional Similarity, Phase Transitions and Hierarchical Clustering," Working Notes of the AAAI Fall Symposium on Probabilistic Approaches to Natural Language, pp. 108-112, 1992.
- [16] F. Pereira, N. Tishby, and L. Lee, "Distributional Clustering of English Word," ACL, 1993.
- [17] D. Petitpierre, S. Krauwer, D. Arnold, and G. B. Varile, "A Model for Preference," ACL, 1987.
- [18] P. S. Resnik, Selection and Information : A Class Based Approach to Lexical Relationships, Ph.D. Thesis, Univ. of Pennsylvania, 1993.
- [19] K. Y. Su and J. S. Chang, "Semantics and Syntactic Aspects of Score Function," COLING, 1988.
- [20] E. Sumida and H. Iida, "Experiments and Prospects of Example-based Machine Translation," ACL, 1991.
- [21] S. Wermter, "Intergration of Semantic and Syntactic Constraints for Structural Noun Phrase Disambiguation," IJCAI, 1989
- [22] Y. Wilks, "Right Attachment and Preference Semantics," ACL, 1985..
- [23] D. Yarowsky, "Word-Sense Disambiguation Using Statistical Models of Roget's Categories Trained on large Corpora," COLING, 1992.

[24] D. Yarowsky, "Unsupervised Word Sense Disambiguation rivaling Supervised Methods," ACL, pp. 189-196, 1995.

[25] 강원석, 신경망을 이용한 영한 기계번역에서의 전치사구 의미해석, 한국과학기술원 전산학과 박사학위 논문, 1995.

[26] 친기석, 국어의 동작 동사와 상대 동사의 체계 연구, 형설출판사, 1984.

1144. 도구 : 용기, 식기, 문구, 악기, 무기
 1145. 탈 것 : 기차, 비행기, 배, 차
 1146. 물품 : 화폐, 책

12. 구체적 장소
 121. 지역 : 서울, 한국, 아시아
 122. 자연
 1221. 천체 : 지구, 태양, 달, 흑성
 1222. 지형 : 호수, 강, 바다, 섬, 산, 해안, 광산
 123. 인공적 장소 : 공원, 학교, 우체국, 회사, 고속도로, 연구소

13. 구체적인 조직 : 회사, 학교, 연구소, 위원회

2. 추상

21. 추상물

211. 자연 추상물 : 빛, 소리, 열, 바람
 212. 인간 활동에 의한 추상물
 2121. 감정적 추상물 : 감정, 애정, 공포
 2122. 감각적 추상물 : 감각, 시각, 촉각
 2123. 이성적 추상물 : 자아, 신념, 소망, 의지, 정신
 2124. 지적 추상물
 21241. 법, 제도, 규칙
 21242. 힘, 권력
 21243. 학문 : 수학, 국어, 귀납법
 21244. 언어
 21245. 사고내용, 지식, 아이디어
 21246. 문화, 풍습
 21247. 종교 : 기독교, 불교

22. 추상사

221. 활동

2211. 행위 : 이동, 결혼, 접촉, 분리, 존재
 2212. 감정적 활동 : 증오, 사랑, 기대
 2213. 사고적(이성적) 활동 : 판단, 생각, 이해
 2214. 감각적 행위 : 시청

222. 현상

2221. 자연 현상 : 일식, 열대야
 2222. 사회 현상 : 혁명, 불황, 데모, 농성
 2223. 물리 현상
 2224. 생리 현상 : 발육, 호흡, 비탄, 소화

23. 속성

231. 장소 : 내부, 상, 하, 중심, 주변, 동, 북
 232. 시간 : 오늘, 시기, 가을, 계절, 미래
 233. 속성명 : 장소와 시간 이외의 속성을 나타내는 단어
 2331. 성질 : 성격, 기질, 맛, 냄새

부 록

A. 명사의 의미 분류 체계

1. 구체

11. 구체물

111. 생물 : 생명이 있는 구체물
 1111. 인간 : 남, 여, 사람, 어린이, 어머니, 과장
 1112. 동물(인간 이외의 동물) : 개, 코끼리, 곤충
 1113. 식물 : 개나리, 소나무, 분재

112. 생물의 부분 및 파생물
 1121. 동물(인간 포함)의 일부 및 파생물 : 손, 발, 심장, 알
 1122. 식물의 일부 및 파생물 : 잎, 줄기, 뿌리, 모근

113. 자연물 : 무생물로 존재하는 자연물
 1131. 물질
 11311. 고체 : 금, 은, 바위
 11312. 액체 : 물, 수은, 석유
 11313. 기체 : 공기, 수증기, 질소, 산소
 1132. 기상 현상물 : 구름, 눈, 비

114. 인공물
 1141. 재료, 자재 : 종이, 섬유, 목재, 벽돌
 1142. 의류 : 천, 의복, 모자, 띠
 1143. 식료 : 쌀, 과자, 음료, 과일

- 2332. 상태 : 상황, 상태
- 2333. 외관 : 형상, 색, 모양
- 2334. 류, 형 : 형식, 종류
- 2335. 수량 : 강우량, 신장, 길이
- 2336. 관계 : 유무
- 2337. 기능 : 기능, 성능, 능력
- 234. 속성치 : 속성의 정도를 나타내는 단어
 - 2341. 성질 : 신속, 침착
 - 2342. 상태 : 위험, 안전
 - 2343. 외관 : 시각, 보라색, 원형
 - 2344. 류, 형 : 집기류, 고딕형, 중세식
 - 2345. 수량 : 10개, 25세, 대용량
 - 2346. 관계 : 밀접, 상이, 유사
 - 2347. 기능 : 고성능, 다기능

B. 동사의 의미 분류 체계

1. 동작 동사

- 11. 수여 의미 : 행위자가 수혜자에게 대상물을 전달한다.
 - 111. 대상물의 제공 : 주다, 보내다, 전하다, 제공하다
 - 112. 소유 이전 : 팔다, 주다, 임대하다, 매매하다
 - 113. 발화 전달 : 말하다, 묻다, 설명하다, 명령하다, 이야기하다
 - 114. 수혜자의 격하 : 모욕하다, 욕하다, 비난하다, 조롱하다
 - 115. 행위자의 시현 : 연주하다, 춤추다
 - 116. 수혜자에 대한 기여 : 부양하다, 보호하다, 도와주다
 - 117. 요구와 결정 : 허락하다
- 12. 소유 의미 : 어떤 대상의 소유주와는 관계없이 소유권이 소유주에게 이양된다.
 - 121. 단순 소유 : 가지다, 소유하다, 보유하다
 - 122. 전체 소유 : 연다, 차지하다, 빌리다, 수입하다, 매입하다
 - 123. 도구 조건 : 이해하다, 탐문하다, 탐지하다
 - 124. 내재 활동 : 주의하다, 조심하다, 염려하다, 생각하다, 우려하다
 - 125. 체내 공급 : 먹다, 마시다, 삼키다, 식사하다, 음주하다
 - 126. 수혜자 손실 : 다치다, 격다, 부상하다, 패배하다, 앓다, 망하다
 - 127. 추상 소유 : 예상하다, 예측하다, 원하다, 바라다, 밀다, 생각하다

- 13. 이동 의미 : 개체의 공간적인 위치가 변화된다
 - 131. 생명체의 이동 : 가다, 오다, 왕래하다, 이동하다, 옮기다, 이전하다
 - 132. 무생명체의 이동 : 돌다, 흐르다, 뜨다, 자전하다
- 14. 부정 의미
 - 141. 결과의 반대 : 부정하다, 부인하다, 반대하다
 - 142. 규칙의 위반 : 위반하다, 위약하다
- 15. 비교 의미 : 행위자가 대상과 대상, 사실과 사실을 상호 비교하는 의미. 분간하다, 비교하다, 알보다, 차별하다, 일치하다
- 16. 상접 의미 : 같은 자격을 가진 복수 명사구를 요구한다.
 - 161. 의지 일치 : 약속하다, 연애하다, 결혼하다
 - 162. 의지 이탈 : 싸우다, 언쟁하다, 전쟁하다
- 17. 조종 의미 : 행위자가 어떤 대상이나 주체를 조종함으로써 그 대상이나 주체를 변화시키는 기능을 형성한다.
 - 171. 대상 분할 : 파괴하다, 분할하다, 분해하다, 썰다, 분산하다
 - 172. 대상 형성과 종합 : 출판하다, 제조하다, 만들다, 보완하다, 보충하다
- 18. 보류 의미 : 행위자가 행위를 일시적으로 보류하다.
 - 침묵하다, 중지하다, 휴교하다, 쉬다, 결식하다
- 19. 가변 의미 : 어떤 대상이나 사건의 진전이 점진적으로 변이되어 가다.
 - 191. 대상 증대 : 살찌다, 팽창하다, 성장하다, 증가하다, 증대하다
 - 191. 대상 삭감 : 쇠퇴하다, 사망하다, 산화하다, 소모하다, 부패하다, 증발하다

2. 상태 동사

- 21. 도량 의미
 - 211. 기본 도량 : 장단 표시, 광협 표시, 경중 표시, 속도 표시, 수량 표시
 - 212. 부차 도량 : 강세 표시, 선조 표시, 시각 표시
- 22. 평결 의미
 - 221. 추론 의미 : 판별 의미, 능력 의미, 미추 의미, 존재 의미
 - 222. 감각 의미 : 청각 의미, 후각 의미, 미각 의미, 촉각 의미, 온각 의미, 통각 의미, 인체감각 의미, 감정 의미

C. 명사의 의미 속성 및 변환 함수

명사의 의미 속성	해당 의미 분류 항목 번호
1. 구체적 개념	1(...)
2. 추상적 개념	2(...)
3. 인공적 개념	114(...), 123, 13, 2124(...)
4. 자연적 개념	113(...), 122(...), 211, 2221
5. 생물	111(...)
6. 동물	1112
7. 식물	1113
8. 사람	1111
9. 조직	13
10. 가상 현상물	1133
11. 일부 또는 파생물	112(...)
12. 재료, 물질	1132(...), 1141
13. 도구	114(...)
14. 인간 활동의 추상물	212(...)
15. 지적 추상물	2124(...)
16. 장소	12(...), 231
17. 시간	232
18. 상황	2332
19. 활동	221(...)
20. 행위	2211
21. 감정적 활동	2212
22. 사고적(이성적) 활동	2213
23. 현상	222(...)
24. 속성	23(...)
25. 수량	2335
26. 성질 및 상태	2331, 2332
27. 외관 및 유형	2333, 2334
28. 관계	2336
29. 속성명	233(...)
30. 속성치	234(...)

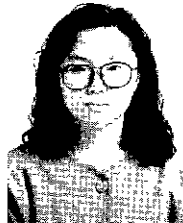
* (...)는 해당 하위 범주의 모든 개념을 포함함을 의미한다.

김 길 창



1963 미국 미시간 대학 전기 공
학과 학사
1966 미국 텍사스 오스틴 대학
수학과 석사
1969 미국 텍사스 오스틴 대학
수학과 박사
1969~70 미우주항공국 연구원
1970~71 미공군사령부 연구원
1972~73 IBM Watson 연구소
초빙연구원
1986~87 일본 NEC C-&C 연
구소 초빙연구원
1973~현재 한국과학기술원 전산학과 교수

조 정 미



1991 이화여자대학교 자연과학
대학 전산학과 학사
1993 한국과학기술원 전산학과
석사
1993~현재 한국과학기술원 전
산학과 박사과정